

京都大学大学院工学研究科

社会基盤工学専攻修士論文

令和8年2月

Master's Thesis

Department of Civil and Earth Resources Engineering

Graduate School of Engineering

Kyoto University

February 2026



---

**水面屈折モデルを導入した  
Gaussian Splattingによる河川浅水域の三次元計測**

**Refraction-Aware Gaussian Splatting  
for Shallow Water Bathymetry from UAV Imagery**

京都大学大学院 工学研究科  
社会基盤工学専攻 空間情報学講座  
宇野 大輝

## 論文要旨

ACRS か ISPRS の内容を翻訳しているだけ。最後にどこまでできたか踏まえて書く。内容も増えるはず。

水中の 3 次元地形の計測は水深測量 (Bathymetry) と呼ばれている。

浅水域においても、河床・海底の水深測量は、地形変動のモニタリング、ハザードシミュレーション、および水生生物の生息環境評価において極めて重要である。近年、無人航空機 (Unmanned Aerial Vehicle: UAV) を用いた写真測量 (Photogrammetry) は、広範囲を効率的に調査する手法として注目されている。しかし、空中から水底を撮影する際、水面で発生する光の屈折が、従来の写真測量における幾何学的仮定（共線性条件）を根本的に破綻させるという課題がある。既存の手法は、屈折を完全に正確に物理的正確性を持ってモデル化しない経験的な補正や反復的な後処理に依存するか、あるいは説明可能性を欠くブラックボックス的な深層学習モデルを用いるものが多く、形状の幾何学的忠実性と外観の写実性を両立させることは困難であった。

本論文では、この課題を解決するために、物理的に忠実な二媒質屈折モデルを再構成パイプラインに直接組み込んだ “Refraction-aware 3D Gaussian Splatting” を提案する。本手法の核心的な貢献は、水中の真の位置にある 3D Gaussian を、航空画像上の見かけの位置へと解析的にマッピングする微分可能なパラメータ変換の導入にある。これにより、標準的な 3D Gaussian Splatting の柔軟なフレームワークを維持しつつ、屈折あり画像からの密な 3 次元形状と詳細なテクスチャ情報の復元を実現した。

評価実験では、屈折以外の光学的要因を排除し厳密な検証を行うため、物理ベースのレイトレーシングにより生成された河床のシミュレーションデータセットを用いた。その結果、水深 10 m のスケールにおいて許容誤差 10 cm とした場合の幾何学的 F1 スコアは 96% を達成した。さらに、新規視点合成 (Novel View Synthesis) においては、PSNR 25.9 dB、SSIM 0.93 という詳細な屈折なし画像の推定を達成した。提案手法により、平坦な水面条件下において、河川、湖沼、沿岸域の低コストかつ高頻度な 3D モニタリングが可能となり、水域リモートセンシング分野に新たな方法論的基盤を提供する。

# 目次

|   |           |
|---|-----------|
| <b>第1章 研究概要</b>                                       | <b>1</b>  |
| 1.1 研究背景  | 1         |
| 1.2 研究課題  | 2         |
| 1.3 研究の目的と貢献  | 3         |
| 1.4 論文構成  | 4         |
| <b>第2章 研究背景と関連研究</b>                                  | <b>5</b>  |
| 2.1 研究背景: 浅水域における水深測量の重要性                             | 5         |
| 2.2 水深測量の諸手法とその課題                                     | 8         |
| 2.2.1 音響測深  | 8         |
| 2.2.2 航空レーザ測深 (Airborne LiDAR Bathymetry)             | 10        |
| 2.2.3 Spectrally Derived Bathymetry                   | 10        |
| 2.2.4 空中写真測量 (Photogrammetric Bathymetry)             | 12        |
| 2.3 水深写真測量の諸手法  | 13        |
| <b>第3章 背景理論</b>                                       | <b>15</b> |
| 3.1 画像に基づく3次元復元 (3D Reconstruction from Images)       | 15        |
| 3.1.1 Structure from Motion                           | 15        |
| 3.1.2 Multi-View Stereo                               | 16        |
| 3.1.3 PatchMatch Stereo                               | 18        |
| 3.1.4 Feed Forward 3D Reconstruction                  | 20        |
| 3.2 Inverse Rendering as 3D Reconstruction            | 21        |
| 3.2.1 Analysis-by-Synthesis and Novel View Synthesis  | 22        |
| 3.3 Gaussian Splatting                                | 25        |
| 3.3.1 2D Gaussian Splatting                           | 30        |
| 3.4 届折を扱った関連手法  | 32        |
| <b>第4章 提案手法</b>                                       | <b>33</b> |
| 4.1 Position Correction (位置補正)                        | 33        |
| 4.2 3D Covariance $\Sigma^{3D}$ Correction (3D 共分散補正) | 35        |

|  |           |
|--|-----------|
| 4.2.1 Covariance Correction for 3DGS . . . . . | 36        |
| 4.2.2 Covariance Correction for 2DGS . . . . . | 37        |
| 4.3 Color Correction (色補正) . . . . .           | 37        |
| <b>第5章 Dataset</b>                             | <b>40</b> |
| 5.1 Simulated Dataset . . . . .                | 40        |
| <b>第6章 Dataset</b>                             | <b>42</b> |
| 6.1 Field Dataset . . . . .                    | 42        |
| 6.1.1 Study Area . . . . .                     | 42        |
| 6.1.2 UAV Platform . . . . .                   | 43        |
| 6.1.3 Data Acquisition . . . . .               | 44        |
| <b>第7章 Experiments</b>                         | <b>45</b> |
| 7.1 Workflow . . . . .                         | 45        |
| 7.1.1 データ収集とセマンティック・セグメンテーション . . . . .        | 45        |
| <b>第8章 APPENDIX</b>                            | <b>48</b> |
| 8.1 見かけの深度 (Apparent Depth) の導出 . . . . .      | 48        |
| <b>参考文献</b> . . . . .                          | <b>50</b> |

# TODO

|  |   |
|--|---|
| ■ ACRS か ISPRS の内容を翻訳しているだけ。最後にどこまでできたか踏まえて書く。内容も増えるはず。 . . . . .  | 2 |
| ■ 全体的に説明が簡潔すぎて、ボリューム感が足りない。エコトーンなど、研究の立ち位置を追加。GS に関しても増強。Intro と Prelim をきれいに書いて、方法までやってから書くべし。 . . . . .  | 1 |
| ■ 「」や () の全角、半角の不統一は後でまとめて置換する . . . . .   | 1 |
| ■ 浅水域が存在量的にも重要だというふうにも見えるのですが、なぜこれまで測量が進んでこなかったのかをうまく説明できる流れにするのは難しそうなので、どちらかというと端っこでこれまで見逃されてきた、後回しにされてきたけれど、実は水と陸のエコトーンの部分にあたって生態的にも人の資源や観光利用において重要な場所であるというふうなことは書いていけないでしょうか。浅水域の定義みたいなものがあればいいと思いました。深さの範囲とか、川岸、湿地、砂浜、干潟などがそういう場所にあたると思います。浅水域の重要性をイメージできる画像もあれば。 . . . . . | 1 |
| ■ エコトーンの文脈でより詳細に。まず、20intro の研究の意義でしっかり書いてから要約的にまとめる . . . . .   | 1 |
| ■ ここは、20intro の関連研究で詳細に説明 . . . . .  | 1 |
| ■ +水中や水面に生物・無機物の障害物が多くて航行制御しにくすことなども . . . . .   | 1 |
| ■ こないで Ipad で読んだ、PB レンダリングの記事で分かりやすくまとめられていた。図を引用しても良いと思う。ちなみに <a href="https://imadr.me/pbr/">https://imadr.me/pbr/</a> ちなみに Geometrical Optics に屈折は含まれるためこの表現は適切ではない . . . . .   | 2 |
| ■ Yudha さんの手法が回帰 . . . . .   | 2 |
| ■ Survey 論文でも引用しとく . . . . .   | 2 |
| ■ GS の簡潔な説明。Explicit で高速。編集可能性や解釈可能度が高い。詳細は 42prelimGS で詳細に説明。ラスタライズで屈折を表現するのは難しいが、うまくやったぞ。 . . . . .  | 3 |
| ■ これまでどんな問題に導入されてきたか、浅水域の写真測量の気体-液体屈折においてなぜ有効だと思われるか、イントロの中でも簡単に説明ができればいいと思いました。見る角度によって変化するものに対応とか、点を確率的なものに考えることで 0-1 の離散的ではなくより連続的な値が求められる?など。それが、結果としてはどうだったか、最後に考察で議論もしてもらいたいところです。 . . . . .   | 3 |

|  |    |
|--|----|
| ■ 3 DGS に取り組んだのはなぜか、浅水域の特徴、屈折のモデルに関連付けて説明できるとい。例えば、浅水域では地形とともに多様な底質が生物の生息に重要で、溝や多孔質な構造、色の判定も重要で、3 DGS はそれが特異とか、 . . . . .                        | 3  |
| ■ 日本語版の図を作成 . . . . .  | 3  |
| ■ + Field Work と実環境での検証、としてわけても良い . . . . .  | 4  |
| ■ 浅水域を具体的に定義する文章を挿入。水底が視認可能な水域で、現実的には透明度が高く水面が穏やかな水域。具体的には、淡水では 2m、海水では 5m 程度が現実的。 . . . . .   | 5  |
| ■ 植生 = フォトリアルな見た目から判別というニュアンスを含め GS の NVS タスクの有用性を後に説明したい . . . . .  | 8  |
| ■ 浅水域で重要な色やテキスチャーによる底質の判断にはレーザーは不向き . . . . .  | 10 |
| Figure: 各手法の Pros & Cons を Table にする。手法、測定可能深度、空間解像度、浅水域での効率、コスト、頻度、安全性、備考 . . . . .   | 10 |
| ■ SDB に関して特化した Survey 論文があったように思える . . . . .   | 10 |
| ■ Lyzenga 法や Stumpf 法などの引用 . . . . .   | 12 |
| Figure: 図で屈折が既存の光の直進性や Pinhole モデルを成立させなくことを示す . . . . .   | 13 |
| ■ 項目のタイトルが英語と日本語が混在。日本語がある場合はなるべくした方がよいか . . . . .   | 15 |
| ■ 無駄に力入れて書きすぎ。MVS とかほとんど本題じゃない。むしろ SfM を頑張って記述すべきだった。(いい勉強になったが)、Appendix に回した方がいいげ . . . . .  | 16 |
| ■ cite では、現在の設定では名前が表示されずただ、[8] のように番号のみになるため、別の方法や引用のフォーマットの設定を考える . . . . .  | 17 |
| Figure: Rendering と Inverse Rendering を表す図 . . . . .   | 22 |
| ■ いらないと思う . . . . .  | 22 |
| Figure: NeRF と 3DGS の図。レイトレーシングとラスタライズの違いを表すイメージ。Survey 論文に合った。 . . . . .  | 23 |
| Figure: NeRF の MLP の構造も図で載せたい。引用できる . . . . .  | 24 |
| Figure: NeRF と 3DGS の分類図。レンダリング速度と、データサイズの 2D プロット。Barron さんのレクチャー動画から。学習速度も入れれば、2 次元に収まらない。2 個いるかな? できれば、MVS とかも含めて示せればいい。それだったら表が美しい . . . . . | 25 |
| ■ 自動運転 . . . . .   | 26 |
| ■ MCMC の説明も。?。これらのヒューリスティックスはデータセットや初期値に対して依存するため、それを克服するような MCMC といった手法も考案されている。 . . . . .  | 28 |
| ■ ローパスフィルタまでは説明しなくても良い。(3DGS でも説明していないから)。MipSplatting 説明しないといけない . . . . .  | 31 |
| Figure: 依然、木津川で再構成した際のアーティストを示す . . . . .  | 31 |
| ■ メッシュの抽出には TSDF Fusion を用いている。 . . . . .  | 32 |

|   |    |
|---|----|
| ■ 今までの説明から、なぜ屈折を考慮した見た目(Photometric)を再現することで、屈折の影響を除去した正当な3次元形状を復元することが可能となるか。Analysis-by-Synthesisの観点から説明する。   | 33 |
| ■ どちらを入射角、屈折角と呼ぶか悩ましい。Ray-Tracing的に視線をベースに考えると空気中が入射角だし、実際のRadianceを考えると水中が入射角である   | 34 |
| ■ ACRSで誤っていた数式を修正済  | 34 |
| ■ Implementationは後述だが、収束の速さからNewtonを使用  | 34 |
| ■ 詳細な導出はAppendixで。ACRSで記述しているのでそれを翻訳  | 35 |
| ■ 実装は後述だが、中心差分法でなく、陰関数定理を用いた。そのため、水面高さや屈折率に対しても勾配を流し、屈折率やカメラ位置も最適化対象にすることができる   | 35 |
| ■ 3DGSに対する補正と2DGSに対する補正双方を描く  | 35 |
| Figure: それぞれのCorrectionを適用した場合、屈折無しで学習したモデルがどのように真値の屈折あり画像に近づいていくか見せる。Lake1とかを使うといいかも。  | 35 |
| ■ $J_{app}$ の導出をAppendixに載せたい。実装でも結局明示的に算出しているから  | 36 |
| ■ Color Correctionというよりも、‘Radiometric Correction’という方が適切でニュアンスが伝わりやすいと思い変えた   | 37 |
| ■ Blenderでは、どのRGB補正が使われ画像が出力されるか確認  | 37 |
| ■ * 実際の取得画像を示すべき* 新たなBlenderデータセットを完成させ、10種類ほどの多様なデータセットで複合的な検証を行ったかったが、  | 40 |
| ■ まだArea3を使用するか決めていない。2が環境が異なる(大きな礫が多い河床)だが、RTK情報がないため厳しい。3は試したいが、影とそれ以外の輝度差の大きい条件  | 42 |
| ■ *ここでは、撮影エリアでの特徴に終始するぞ。*水面には水草が浮いているぞ。*本流ではなく、水たまりになっているSecond Channelにおいて撮影したぞ。*最大深度は1.5mほどだと推定されるぞ。*Textureも比較的豊富だぞ。*透明度が高く、波も穏やかだぞ。*気温や湿度、天気はこの程度だった。(撮影時14度、湿度60%、快晴)(後に正確に調べ直す) | 42 |
| Figure: 撮影エリアの概要GISを使って、論文らしい格好いい図を作る。1だけでなく、3でもデータを作りたいから、こだわるのはまた後で。  | 42 |
| Figure: 撮影機材の概要   | 43 |
| ■ 根拠となる論文を引っ張ってくる   | 43 |
| Figure: 撮影データの例   | 44 |
| ■ Field Studyの具体的な詳細は前章のDatasetで  | 45 |
| ■ 現在Codeも公開されているが、正直いい実装には思えなかった。   | 46 |
| Figure: 実際のマスク作成例。定量も含めて示す  | 46 |

- Automated: Manual labeling for hundreds images is time consuming ! Robustness: robustly across diverse environments. Zero-Shot: Fine-tuning models for each new environment is also time consuming impractical ! Controllability : flexible specification using natural language. “I want to remove those ‘ship and shadow’ !” . . . . . 47
  - Dino の説明は、40prelim の Forward 3D Reconstruction でシたいと思ったがしないかも . . . . . 47

# 第1章 研究概要

全体的に説明が簡潔すぎて、ボリューム感が足りない。エコトーンなど、研究の立ち位置を追加。GSに関しても増強。IntroとPrelimをきれいに書いて、方法までやってから書くべし。

「」や「」の全角、半角の不統一は後でまとめて置換する

## 1.1 研究背景

浅水域が存在量的にも重要だというふうにも見えるのですが、なぜこれまで測量が進んでこなかったのかをうまく説明できる流れにするのは難しそうなので、どちらかというと端っこでこれまで見逃されてきた、後回しにされてきたけれど、実は水と陸のエコトーンの部分にあたって生態的にも人の資源や観光利用において重要な場所であるというふうなことは書いていけないでしょうか。浅水域の定義みたいなものがあればいいと思いました。深さの範囲とか、川岸、湿地、砂浜、干潟などがそういう場所にあたると思います。浅水域の重要性をイメージできる画像もあれば。

エコトーンの文脈でより詳細に。まず、20introの研究の意義でしっかりと書いてから要約的にまとめる

地球表面の大部分を覆う水域、特に沿岸部や河川などの浅水域(Shallow Water)は、人間社会の経済活動、防災、生態系保全において極めて重要な役割を果たしている。河川管理における氾濫原の地形変状把握 [1]、沿岸域管理 [2]、生態系の生息環境評価 [3] など、水底の詳細な地形データを観測する水深測量(Bathymetry)は非常に重要である。しかしながら、これらの領域は従来の測量技術では取得が困難な空白地帯となりがちであった。

ここは、20introの関連研究で詳細に説明

伝統的な船舶搭載型マルチビームソナー(深浅測量・音響測深)は、一定の深度がある海域においては標準的な手法であるが、水深が極めて浅い河川や海岸線付近においては、船舶の座礁リスクなどの航行不可領域の存在により、その運用は著しく制限される。近年では、小型の無人水上艇(Unmanned Surface Vehicle: USV)を用いた水深測量が注目されているが [4,5]、マルチビームの指向角の制限により、浅水域においては走査幅(Swath Width)が狭く、面的な測量を行うには時間的コストが高くなるという、非効率性の問題も生じる。

+水中や水面に生物・無機物の障害物が多くて航行制御しにくうことなども

一方で、航空機搭載レーザ測深(ALB: Airborne LiDAR Bathymetry)は [6]、広域かつ高精度な計測が可能であるが、導入および運用コストが極めて高く、高頻度なモニタリングには向きであるという経済的な障壁が存在する。

こうした背景の中、近年急速に普及したドローンなどの無人航空機(UAV: Unmanned Aerial Vehicle)を用いた写真測量(Photogrammetry)は、低コストかつ高解像度、高頻度なデータ取得が可能であること

から、次世代の浅瀬測量技術として大きな期待を集めている。UAV により空撮された多視点画像から、Structure-from-Motion (SfM) [7]、および Multi-View Stereo (MVS) [8·9] を用いて 3 次元形状を復元するアプローチは、陸部においては既に広い用途で実用化されている [10–12]。これらの技術を水中に適用する場合、その手法は空中からの水深写真測量 (Photogrammetric Bathymetry) と呼ばれ、空中から撮像した多視点画像からの水中の三次元再構成問題と捉えることができる。水深写真測量には、光の反射や、水中での光の散乱・吸収による減衰、波による被写体の歪みなどの課題が存在するが、中でも最も根本的に重要な課題として取り組まれてきたのが光の屈折 (Refraction) である。

## 1.2 研究課題

既存の SfM・MVS アルゴリズムの大部分は、幾何光学 (Geometrical Optics) を前提としている。

こないで Ipad で読んだ、PB レンダリングの記事で分かりやすくまとめられていた。図を引用しても良いと思う。ちなみに <https://imadr.me/pbr/> ちなみに Geometrical Optics に屈折は含まれるためこの表現は適切ではない

すなわち、撮像 (Image Sensing) のプロセスにおいて、観測対象となる被写体から発せられる光は、被写体からカメラ中心まで直進することを仮定する。しかし、UAV による水中撮影においては、光は水中から空气中へ進む際に、水面と異なる媒質の境界でスネルの法則 (Snell's Law) に従って屈折する。この物理現象によって、カメラから見た被写体の「見かけの位置」(Apparent Appearance) は、実際の位置よりも浅く、近く、歪ませる。屈折の影響を無視し、既存の SfM・MVS アルゴリズムを適用する場合、水深が実際よりも浅く評価される。これに対処するために、従来は SfM・MVS の出力結果に屈折率に基づく補正を適用する手法 [13·14] や、点群とカメラフレームの位置関係から推定する手法 [15·16]、手動で計測した数カ所の真値をもとに出力結果の補正率を回帰で決定する手法

Yudha さんなどの手法が回帰

が提案してきた。こうした手法は、屈折の影響を補正する一方で、屈折の物理的特性を完全にモデル化しているわけではなく、視線角度依存性や多視点間の整合性を厳密に扱えないため、幾何学的精度には限界があった。直近では、[17] が、屈折の物理的特性を直接 SfM の最適化に組み込むことで、より高精度な再構成を実現する手法を提案している。しかし、この手法では、疎な出力結果を補うために、USV など高いコストを要する計測機器を用いた測量結果を用いた深層学習手法で補間する [18] 必要があり、学習データ不足とフィールド依存性という問題がある。

さらに、近年のコンピュータビジョン (Computer Vision)・コンピュータグラフィックス (Computer Graphics) の領域では、Neural Radiance Fields (NeRF)[19] や 3D Gaussian Splatting (3DGS) [20] といった、微分可能なレンダリング (Differentiable Rendering) を用いた新たな 3 次元表現・再構成手法が登場している。これらは任意視点における写真のようなリアルな新規視点合成 (Novel View Synthesis:NVS)において卓越した性能を示し、照明依存性、時間軸方向へ拡張、幾何情報の抽出といった多種多様な課題を克服するよう、日進月歩の進化を遂げている

Survey 論文でも引用しつく

。特に、NeRF のような陰的三次元表現 (Implicit Representation) は計算コストが高く、幾何的な

走査や解釈が容易ではない一方、3DGS は明示的な三次元 Gaussian 点群表現 (Explicit Representation)を持ち、高速かつ直接的な形状操作が可能であるという利点を持つが、空気中から観測する水中屈折を考慮した定式化は未だ十分になされていない。

### 1.3 研究の目的と貢献

GS の簡潔な説明。Explicit で高速。編集可能性や解釈可能度が高い。詳細は 42prelimGS で詳細に説明。ラスタライズで屈折を表現するのは難しいが、うまくやったぞ。

これまでどんな問題に導入してきたか、浅水域の写真測量の気体-液体屈折においてなぜ有効だと思われるか、イントロの中でも簡単に説明ができればいいと思いました。見る角度によって変化するものに対応とか、点を確率的なものに考えることで 0-1 の離散的ではなくより連続的な値が求められる?など。それが、結果としてはどうだったか、最後に考察で議論もしてもらいたいところです。

3 DGS に取り組んだのはなぜか、浅水域の特徴、屈折のモデルに関連付けて説明できるとよい。例えば、浅水域では地形とともに多様な底質が生物の生息に重要で、溝や多孔質な構造、色の判定も重要で、3 DGS はそれが特異とか、

以上の背景から、本修士論文では、UAV 空撮画像からの水中 3 次元復元において、物理的な屈折モデルを Gaussian Splatting(GS) のパイプラインに直接統合することで、幾何学的正確性と写実的な外観再現を両立させる新たな枠組み (Refractive-Aware Gaussian Splatting) を提案・実証する。本手法の核心的な貢献は、水中の真の位置にある 3D Gaussian を、UAV 空撮画像中の見かけの位置へと解析的にマッピングするパラメータ変換にある。この変換は、微分可能であるたり、GS の最適化過程に直接組み込むことで、屈折を含む入力画像から直接、屈折のない三次元シーン (3D Scene) を推定し、屈折の影響を排除した密な 3 次元形状と詳細なテクスチャ情報の復元を実現する。

日本語版の図を作成

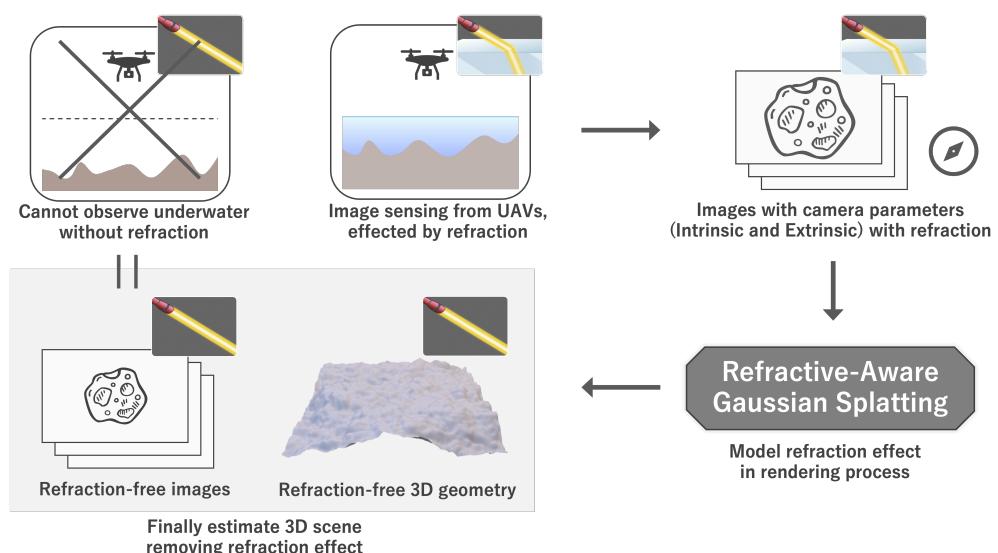


図 1-1: 空気中からの水深写真測量のタスクの概要。空中から水底を撮像したパラメータ既知の多視点画像を入力として、Refractive-Aware Gaussian Splatting を用いることで、屈折の影響を排除した水中の三次元シーンを再構成する。

検証においては、PBR レンダリング (Physically-Based Rendering:PBR) によってシミュレートされた

合成データと実データの双方を用いて検証を行った。合成データでは、不観測地点からの再構成 3D モデルの視覚的品質を指す Peak Signal-to-Noise Ratio (PSNR) が 25 dB を超え、提案手法から抽出した幾何情報では真値との幾何的復元精度を表す F1 スコア (F1-score) で 94% を超える結果を確認した。(合成データは深度 10 m、撮影高度 10 m のスケールである。完全に平面の水面とカメラパラメータ既知を仮定し、屈折の影響のみを考慮し、反射や減衰、波など屈折以外の物理的影響は除外している。F1 スコアは許容誤差 10 cm とした。) 実データにおいても、幾何的誤差 を確認するなど、実用的な空中からの水深写真測量としての方法を確認した。

#### 1.4 論文構成

本論文は導入部である第 1 章を含め、全 6 章で構成される。第 2 章では、水深測量の概要と諸手法といった関連研究に関して述べる。第 3 章では、本研究の基盤となる Gaussian Splatting やコンピュータビジョンの諸手法の学術的背景と理論に関して記述する。第 4 章では、提案手法である Refractive-Aware Gaussian Splatting の詳細な理論と実装に関して記述する。第 5 章では、検証において用いた合成データと実データの双方を用いて、提案手法の性能を定性・定量的に検証する。

+ Field Work と実環境での検証、としてわけても良い

第 6 章では、研究の課題 (Limitation) と今後の課題 (Future Work) に関して述べ、研究の総括を行う。

## 第2章 研究背景と関連研究

浅水域を具体的に定義する文章を挿入。水底が視認可能な水域で、現実的には透明度が高く水面が穏やかな水域。具体的には、淡水では 2m、海水では 5m 程度が現実的。

### 2.1 研究背景: 浅水域における水深測量の重要性

地球表面のおよそ 7 割は、海や河川といった水域によって占められる。近年、日本財団[21]と GEBCO(General Bathymetric Chart of the Oceans)[22]による国際プロジェクト「Seabed 2030」[23]の推進によって、深海域における海底地形図の整備は急速に進展している[24]。対照的に、陸域と水域の境界である「浅水域(Shallow Water)」は、既存の測量技術では効率的な計測が難しく、水路測量分野において‘White Ribbon’と呼ばれるデータの空白地帯となっている[25]。これは、海図や河川図において、陸域の地形データと深部の水深データとの間に挟まれ、有効な測量データが存在しないために白く描かれ、帯状をなすことに由来する。

一方、浅水域は、人間社会の経済活動、防災、そして生態系保全において決定的な役割を果たす領域である。本章では、浅水域における三次元計測の重要性を、河川管理および生態学(エコトーン)の観点から概説し、現状の課題を整理する。

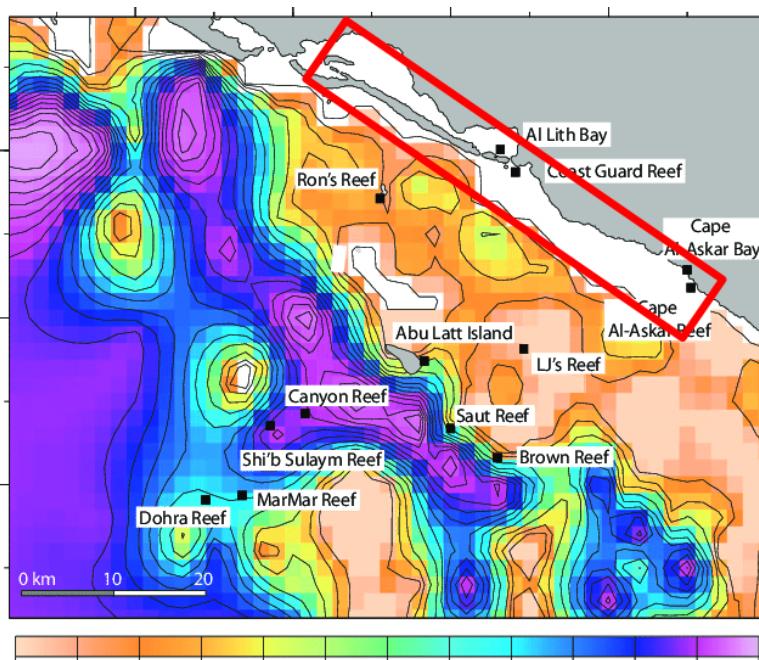


図 2-1: White Ribbon の実例。[26] より引用。[22] の提供する Gridded Bathymetry Data による海洋沿岸の水深測量図。赤枠で囲われた白い領域は浅水域のため測量不能。

## 定期横断測量

河川および沿岸域の浅瀬は、洪水の流下能力を決定する重要な断面であり、津波や高潮に対する第一の防衛線として機能する。特に日本においては、河川法に基づき、一級河川を対象とした定期的な横断測量が義務付けられている [27]。

しかし、現行の定期横断測量には、空間分解能と安全性において問題が存在する。標準的な測量間隔は200メートルピッチとされており、断面間の局所的な地形情報を捉えることができない。また、従来の水部計測は、有人船による音響測深や、測量員が直接川に入りポールで計測する手法に依存している(図2-2)。これらは、水難事故のリスクを伴うだけでなく、測線維持のための慎重な操船や作業をするため、広範囲のデータ取得において著しく時間的効率を欠いている。こうした課題は、少子高齢化による作業員の不足や、河川管理に面的な管理が求められはじめている状況も鑑みて、効率的で高密度な水深測量技術の開発が重要になる。



図2-2: 一級河川縦横断測量の実例。[28]より引用。測量員によるポールによる計測には多大な労力と時間がかかる一方、疎な三次元情報しか得られない。

## 河川地形の時系列変化モニタリング

河川工学の観点からは、浅水域計測には「高解像度」に加え、「高頻度」なモニタリングが求められる。急峻な地形を有し、台風等による豪雨が頻発する日本の河川では、土砂移動が活発であり、出水のたびに局所地形が大きく変化する。そのため、5年に1度といった従来の定期測量サイクルでは、洗掘や堆積といった動的な地形変化を捉えることができない。

この課題は、伝統的河川工法である「聖牛」などの機能評価において顕著に現れる。聖牛は丸太を組み上げた構造物であり(図2-3)、河川砂州部に設置することで流速・流域を制御し、人間社会に寄与する災害の起こりにくい地形形成を導く、「Nature-Based Solutions (NbS)」として、近年再評価されている技術である。砂州などの自然地形が形成されることで、これらは生物の生息地となり、生態系の多様性の基盤ともなる。[1]は、京都府木津川における6年間にわたるモニタリングを通じ、聖牛設置地域一帯

を観測し、聖牛の設置によって地形が大きく変化することを明らかにした(図 2-4)。同研究が示すように、砂州が単調な地形から水路や池が点在する複雑な地形へと遷移する過程を定量的に評価するためには、陸域と水域をシームレスに、かつ高解像度で捉える三次元計測技術が必要不可欠である。

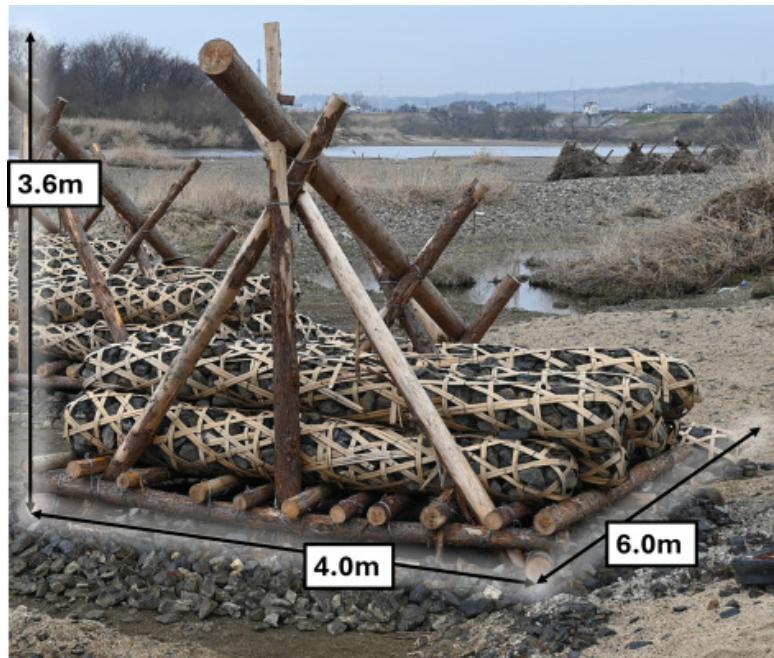


図 2-3: 聖牛の写真。[1] より引用。

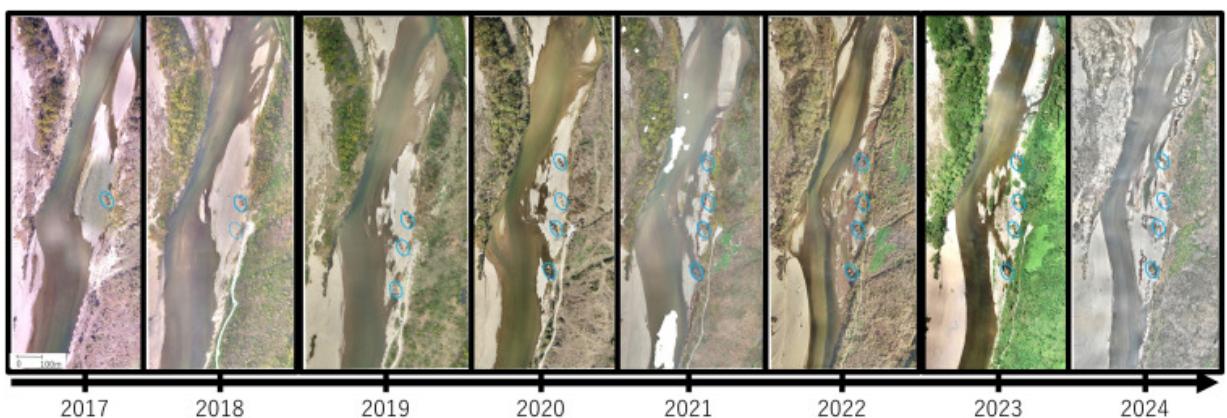


図 2-4: [1] より引用。聖牛による河川地形の時系列変化。

## エコトーン

浅水域の重要性は、防災のみならず生態系保全の観点からも極めて高い。20世紀初頭に提唱された「エコトーン(Ecotone)」は、二つの異なる生態系が接し移行する境界領域を指し [29]、陸域から水域へと移行する浅水域一帯はこれに該当する。この領域は、陸と水双方の環境要因が作用することで、高い生物多様性を有するエッジ効果(Edge Effect)が発現する場として知られる [<empty citation>]。

エコトーンの形状と生態学的プロセスの関係については、多くの研究がなされている。例えば、[30] は乾燥地帯の植生分布が旧河道などの微地形に依存することを示し、[31] は珊瑚礁の三次元構造が生物

侵食速度や多様性に与える影響を報告している。ドローン空中写真測量により、サンゴ礁など浅水域の高解像度三次元モデルと色情報を同時に取得できる。こうしたデータは、底生生物群集の多様性や経時的な変化を定量評価する基盤情報となり、広域マッピングや生息地モニタリングにも有用である [32]。これらの研究は、エコトーンの環境保全において、詳細な三次元地形情報が重要であることを示唆している。

浅水域エコトーンにおける生態系の豊かさは、太陽光の到達に由来する。水生植物や藻類の基礎生産が行われる水深帯は「有光層(Photic Zone)」と呼ばれ、水底まで十分な光が届くことが生物生息の条件となる。一方、写真測量は受動的光学手法(Passive Optical Method)であり、原理的に水底からの反射光を必要とするため、濁度や水深による制約を受ける手法である。一方、受動的リモートセンシング(Passive Remote Sensing)である写真測量では、濁度や水深による制約が少なく、水底が確認可能である必要がある。従って、「生物多様性が高く計測ニーズのある浅水域」は、必然的に「光が水底まで達する写真測量が可能な場所」である可能性が高いと言える。深すぎて光が届かない領域は、従来の音響測深が有効な領域であるが、浅水域エコトーンにおいては、光学的計測は合理的なアプローチとなる。また、エコトーンの理解には水域だけでなく連続する陸域の植生と地形の把握も同時に求められるため、水中のみを計測するUSVなどによる音響測深は適さない。

植生 = フォトリアルな見た目から判別というニュアンスを含め GS の NVS タスクの有用性を後に説明したい

以上の点から、陸域と浅水域を同時に、かつ高密度に計測可能な空中写真測量は、エコトーンの地形計測(Geomorphology)において最適な手法であると位置付けられる。



(a)



(b)

図 2-5: エコトーンの例: 浅水域一帯では、陸域と水域の地形中に多様な植生が見られる。

## 2.2 水深測量の諸手法とその課題

本節では、既存の水深測量技術を概観し、浅水域(Shallow Water)の三次元計測において生じる技術的課題を整理する。

### 2.2.1 音響測深

現在、水深測量(Bathymetry)の標準となっているのは、船舶に搭載したマルチビーム音響測深機(Multibeam Echosounder: MBES)である。MBESは船底から扇状(Fan-shape)に音波を発射し、走査線上の多数の計測点の水深を同時に取得することで、面的な地形図を作成を可能とする。深海域において

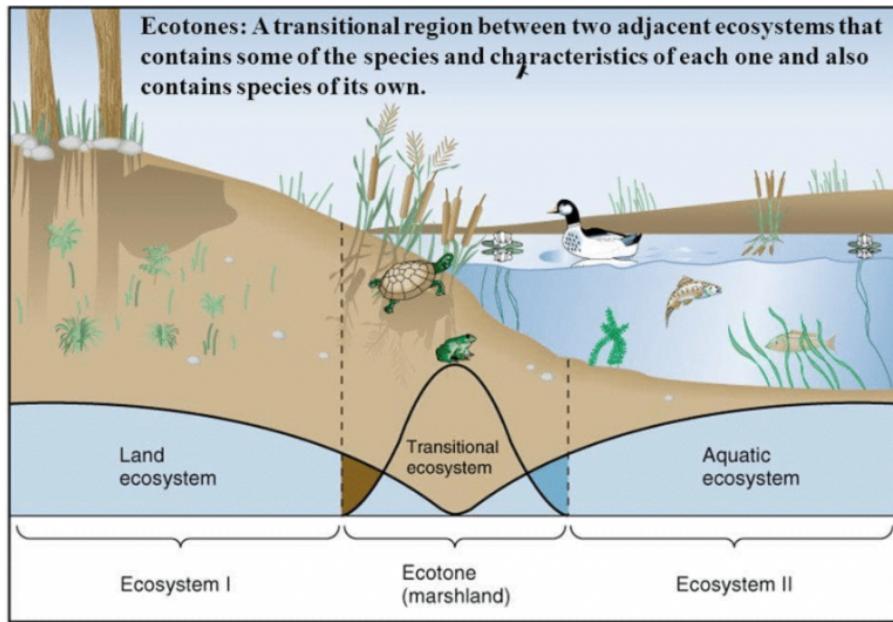


図 2-6: エコトーンの模式図。[33] より引用。陸域と水域の境界にあたる浅水域は、双方の生態系が重なり合う重要な領域である。

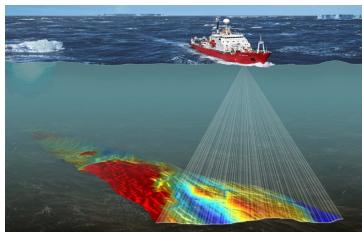
は、一度の航行で数キロメートル幅の海底をスキャンできるため、Seabed 2030 のような大規模な海洋マッピングプロジェクトの中核技術となっている。

しかし、浅水域において MBES の計測効率は大きく低下する。MBES の走査幅 (Swath Width:  $SW$ ) は、幾何学的に水深  $D$  と指向角  $\theta$ (一般に  $120^\circ \sim 150^\circ$ ) に依存して決定されるためである。下式のようにその走査幅は水深に比例する:

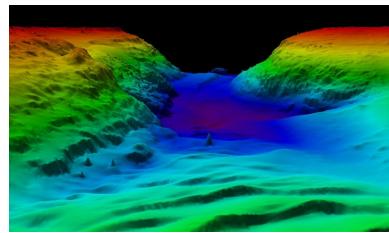
$$SW = 2D \tan\left(\frac{\theta}{2}\right)$$

水深 1000 m であれば 3000 m 以上の幅を一度に計測可能であるが、日本の河川のような水深 1 m ~ 2 m の環境では、走査幅はわずか 3 m ~ 8 m 程度に留まる。したがって、対象水域を網羅するためには、探査船は数十回もの往復(測線)を繰り返す必要があり、時間的・金銭的コストの増大を招く。また、送受波器の直下数十センチメートルは、音波の振動の余韻(リングギング)といった制約により、計測が不可能となる。(例: Teledyne Marine 社製 SeaBat T20-P では?? 深度 50 センチメートル以浅は、計測が不可能である)。

加えて、物理的な制約も存在する。浅水域においては、有人調査船は、座礁のリスクがあるため船体の喫水より浅い場所には侵入できない場合がある。この課題に対し、近年では小型の無人水上艇(Unmanned Surface Vehicle: USV)を用いた測量も研究されている [45]。USV は浅水域へのアクセスを可能にするが、流水環境下での姿勢制御や自己位置推定の難易度が高く、また走査幅の物理的制約(第 2.2.1 小節)は浅水域ではより一層深刻になるため、広範囲を高頻度で計測するには依然としてコストが高い。さらに、音響測深は水域のみに限定されるため、陸域から水域へ連続する地形を単独で計測することは不可能であり、UAV 写真測量等とのデータ統合が別途必要となる。



(a) マルチビームソナーの概念図



(b) マルチビームソナーによる水深測量図



(c) 無人水上艇 (USV) による水深測量例

図 2-7: マルチビームソナーによる水深測量。[4] より引用。

### 2.2.2 航空レーザ測深 (Airborne LiDAR Bathymetry)

航空レーザ測深 (Airborne LiDAR Bathymetry: ALB) は、航空機から水を透過する緑色レーザ (波長 532 nm 付近) を照射し、水面反射と水底反射の時間差から水深を求める技術である [6]。自らエネルギーを照射して観測を行うため、能動的リモートセンシング (Active Remote Sensing) に分類される。濁度などの水質に大きく依存するが、最大で 15 m 程度まで計測可能であり、[34] によると水深 3m 程度までを高密度に測定可能である。

浅水域で重要な色やテキスチャーによる底質の判断にはレーザーは向き

日本国内においても、国土交通省により一級河川の定期横断測量への導入が進められており [35]、陸域と水域をシームレスに、かつ高密度に計測できる点が大きな利点である。

しかし、ALB には本研究が対象とする「浅水域の高頻度モニタリング」において、導入・運用コストの課題を抱えている。有人航空機や大型ドローンに搭載する高出力 LiDAR は極めて高価 (数千万円規模) である。国家規模で行う 5 年に 1 度の定期測量には適しているが、出水のたびに地形変化を追跡するような機動的な運用を、各研究室規模で行うことは経済的に難しい。

Missing figure

各手法の Pros & Cons を Table にする。手法、測定可能深度、空間解像度、浅水域での効率、コスト、頻度、安全性、備考

### 2.2.3 Spectrally Derived Bathymetry

分光水深測量 (Spectrally Derived Bathymetry: SDB) は、Landsat や Sentinel-2 などのマルチスペクトル衛星画像の分光特性を利用し、放射伝達モデルに基づいて水深を推定する手法である [37]。

SDB に関して特化した Survey 論文があったように思える

本手法は受動的リモートセンシング (Passive Remote Sensing) に分類され、広域かつ低コストに水深情報を取得できる利点がある一方で、その推定原理に起因する理論的・実用的な制約が存在する。

SDB の基本原理は、水中における光の指数関数的な減衰に基づく。可視光領域において、水深が深いほど水底からの反射光は水柱による吸収・散乱を受けて減衰する。SDB はこの物理現象を利用し、主

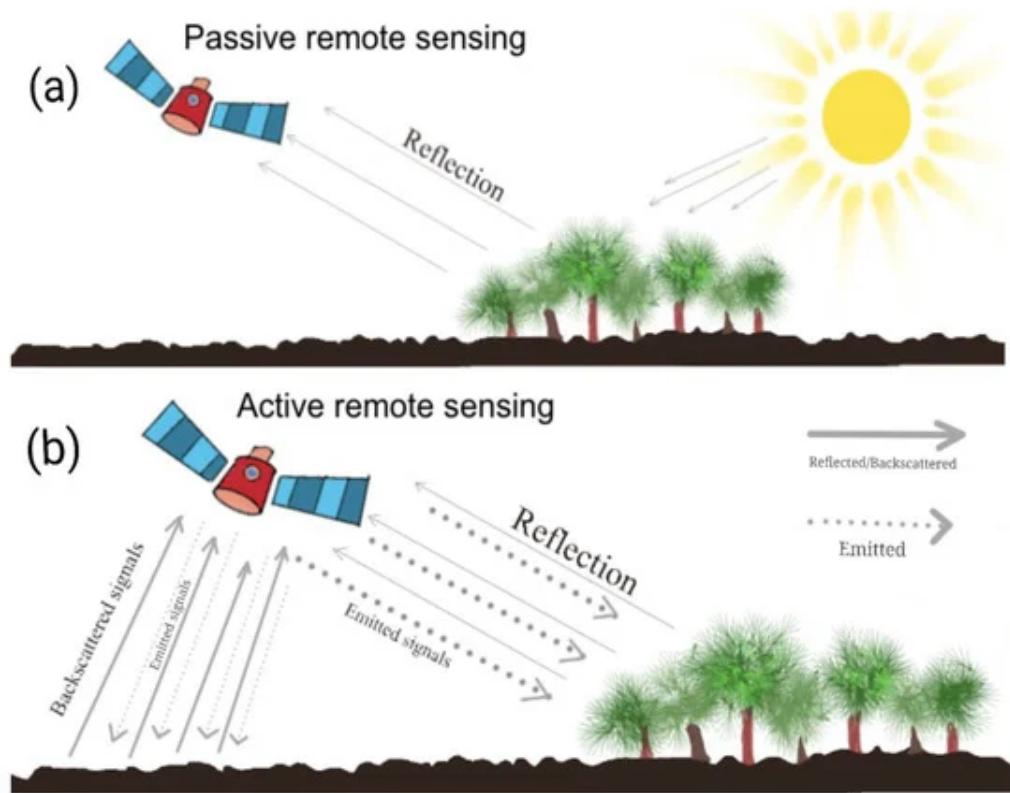


図 2-8: 能動的リモートセンシング(Active Remote Sensing)と受動的リモートセンシング(Passive Remote Sensing)の比較。[36] より引用。(a) 受動的リモートセンシング：センサが外部からの情報を受信する。(b) 能動的リモートセンシング：センサが自ら情報を発信し、その反射を受信する。

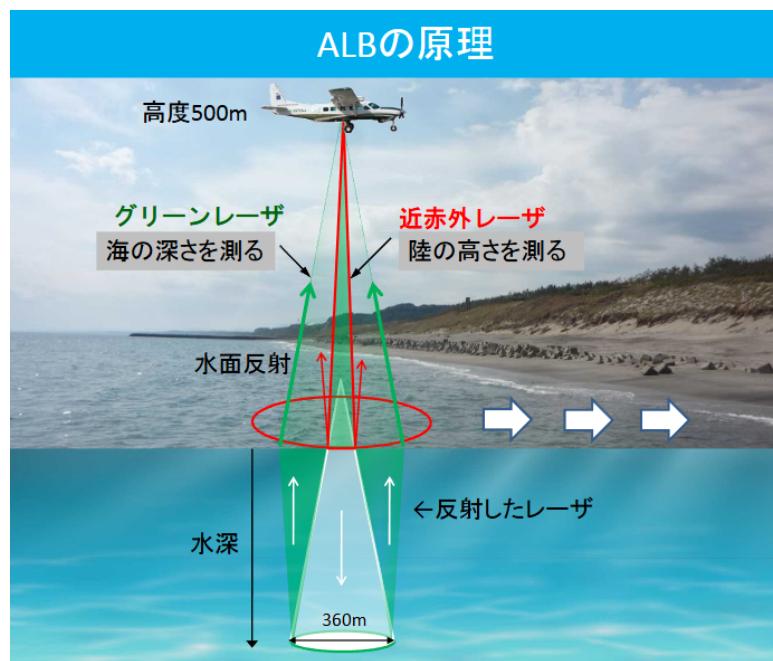


図 2-9: 航空レーザ測深(Airborne LiDAR Bathymetry: ALB)の原理。[34] より引用。水面で反射する近赤外レーザと、水中を透過し水底で反射するグリーンレーザを用い、水底の三次元点群を得る。

に青緑バンド(水中透過率が高い)と赤～近赤外バンド(水中減衰が大きい)の輝度比や、各バンドの反射率の対数線形モデル(Lyzenga法やStumpf法など)を用いて水深を回帰的に推定する。

#### Lyzenga法やStumpf法などの引用

SDBは海洋沿岸部等の広域な浅海域では有効な手法である一方、中小河川環境への適用においては、幾何学的アプローチである写真測量と比較して以下のような限界を有する。

- 放射量依存性と水質・底質の不均一性:** SDBは幾何学的な三次元復元ではなく、観測された放射輝度(Radiance)に基づく放射量的な推定(Radiometric estimation)である。そのため、推定精度は水域に固有の光学的特性(濁度やクロロフィルa濃度)や、水底の底質(砂、礫、植生)によるアルベドの空間的不均一性に強く依存する。
- 経験的モデルと現地データの必要性:** 一般的なバンド比法などの経験的モデルでは、輝度値を水深値へ変換するために現地での実測水深データを用いたキャリブレーション(回帰モデルの補正)が不可欠である。これは「計測なしで水深を得る」というリモートセンシングの利点を部分的に損なうものである。
- 空間解像度とミクセル問題:** Sentinel-2(10 m)やLandsat(30 m)などのオープンデータである衛星画像は、中小河川における三次元情報のニーズに対して解像度が不足する。1ピクセル内に水域、陸域、河畔林が混在する「ミクセル(Mixed Pixel)」問題が発生し、水深推定を阻害する。

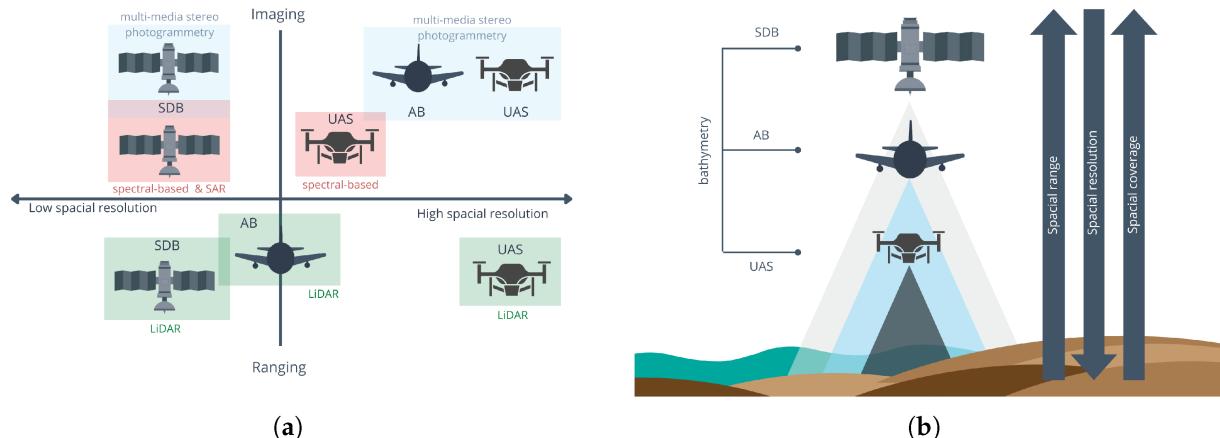


図 2-10: リモートセンシングによる水深測量技術の分類。[38] より引用。(a) 能動的・受動的手法の分類。(b) 各手法の空間分解能とカバレッジの関係。

#### 2.2.4 空中写真測量 (Photogrammetric Bathymetry)

以上の既存手法の課題を踏まえ、近年急速に普及したドローンなどの無人航空機(Unmanned Aerial Vehicle: UAV)用いた写真測量(Photogrammetry)は、低コスト・高解像度・高頻度なデータ取得が可能であることから、浅瀬測量技術においても有効であると考えられる。UAVにより空撮された多視点画像から、詳細を後述する Structure-from-Motion (SfM) [7] および Multi-View Stereo (MVS) [8,9] を用いて3次元形状を復元するアプローチ第3.1節は、陸部においては既に広い用途で実用化されている[10-12]。

この技術を水中に適用する場合、その手法は「空中からの水深写真測量 (Photogrammetric Bathymetry)」と呼ばれ、多視点画像からの水中三次元再構成問題として定式化される。これを実現するためには、光の反射や、水中での光の散乱・吸収による減衰、波による被写体の歪みなどの課題が存在するが、中でも最も根本的で重要な課題として、水面における「光の屈折 (Refraction)」を克服しなければならない。



Missing  
figure

図で屈折が既存の光の直進性や Pinhole モデルを成立させなくことを示す

既存の SfM/MVS アルゴリズムの大部分は、カメラと被写体が同一媒質 (空気中) にあることを前提とし、光が直進するという幾何的な仮定に基づいている。しかし、上空からの水中撮影においては、光は水中から空気中へ進む際に、媒質境界 (水面) でスネルの法則 (Snell's Law) に従って屈折する。この屈折現象により、カメラから観測される被写体の「見かけの位置 (Apparent Position)」は、実際の位置よりも浅く、かつ歪んで観測される。したがって、従来の陸上用アルゴリズムをそのまま適用すると、水深が過小評価され、復元形状が破綻する。本研究では、この屈折を物理的正確に考慮した (Refraction-aware) 新たな三次元再構成手法を導入することで、水面下の高精度な計測を実現する。

## 2.3 水深写真測量の諸手法

<https://www.notion.so/Photometric-Bathymetry-2e002f0efa178010b4f3f365fdea4f29>

以下の表 2-1 に空中からの水深写真測量 (Photogrammetric Bathymetry) における屈折補正手法を比較する。一部は [38] を参考にしている。

表 2-1: 空中からの水深写真測量 (Photogrammetric Bathymetry) における屈折補正手法の比較

| Category         | Reference & Approach                  | Workflow Core Mechanism  | Performance & Limitations   |
|------------------|---------------------------------------|--|---|
| Post-Processing  | Woodget et al. [14]                   | <b>Simple Refraction Correction</b><br>SfM で得られた見かけの水深に対し、水の屈折率 ( $n \approx 1.34$ ) を乗数として適用し、一律に深度を補正する。 | <b>課題:</b> <ul style="list-style-type: none"><li>観測角を考慮せず直下視と仮定するため、画像端部で誤差が増大。</li><li>平坦な水面を仮定。</li></ul>                               |
|                  | BathySfM<br>(Dietrich et al. [16])    | <b>Iterative View Dependant</b><br>SfM 後の点群に対し、カメラ位置から水面を通してレイを追跡し、スネルの法則に基づいて反復的に真の位置を計算する。               | <b>利点:</b> 視野角による影響を考慮可能。<br><b>課題:</b> <ul style="list-style-type: none"><li>水平変位の補正が不十分な場合がある。</li><li>反復計算により処理コスト増。</li></ul>         |
| Image-based      | Agrafiotis et al. [39]                | 画像ベースの屈折補正<br>SfM 処理を行う前に、生画像に対して屈折除去の幾何変換を行い、屈折フリーな画像を生成してから 3D 再構成を行う。                                   | <b>利点:</b> 既存の SfM パイプラインを利用可能。<br><b>課題:</b> <ul style="list-style-type: none"><li>穏やかな水面条件と、海底テクスチャに強く依存。</li></ul>                     |
| Machine Learning | BathyNet<br>(Mandlburger et al. [40]) | <b>Radiometric Depth Estimation</b><br>マルチスペクトル画像の反射輝度と、ALB の教師データを用いて CNN により画素単位の水深を推定。                  | <b>利点:</b> 非線形な深度推定モデルを獲得可能。<br><b>課題:</b> <ul style="list-style-type: none"><li>高品質な教師データ (ALB) が不可欠。</li><li>未学習環境への汎化性能が限定期。</li></ul> |
| Refraction-Aware | R-SfM<br>(Makris et al. [17])         | <b>Refractive Bundle Adjustment</b><br>スネルの法則をバンドル調整 (BA) に組み込み、カメラ外部標定と海底座標を同時最適化。                        | <b>利点:</b> 幾何学的に厳密。カメラと海底の両方を補正。<br><b>課題:</b> <ul style="list-style-type: none"><li>密な点群生成には Deep Learning 依存が残る。</li></ul>              |
|                  | NeRFrac<br>(Zhan et al. [41])         | <b>NeRF with Refractive Surface</b><br>NeRF を拡張し、屈折を伴う非線形なレイ追跡を行うことで 3D 形状を再構成。                            | <b>利点:</b> 波のある水面 (Non-planar) も考慮可能。<br><b>課題:</b> <ul style="list-style-type: none"><li>計算コストが極めて高い。</li></ul>                          |

※ 脚注

# 第3章 背景理論

## 3.1 画像に基づく3次元復元 (3D Reconstruction from Images)

画像を用いた測量は写真測量 (Photogrammetry) と呼ばれ、19世紀半ばに航空写真測量の技術として確立されて以来、長い歴史を持つ。初期の写真測量は、専門的な技能を持つ技術者が専用の測量機器を用いて手動で行うものであったが、画像のデジタル化とイメージセンサの発明により、コンピュータビジョン (Computer Vision, CV) の分野として発展した。コンピュータビジョンは、画像センサから得られた視覚情報を処理し、人間や他の生物と同様に、コンピュータや機械がシーンを理解することを目指す学問分野である。

画像センサ (Image Sensor) が捉える情報は、本質的には3次元シーンから2次元平面への射影 (Projection) である。この過程において、3次元空間の奥行き情報は1次元分欠落し、情報の「縮退」が発生する。3次元復元の主眼は、この失われた次元を幾何学的制約や事前知識 (Priors) を用いて補完し、元のシーンの構造を逆問題として解くことにある。3次元情報は、生物が自己が生きる世界を認識・理解するうえで必須の情報であり、ロボットの自己位置推定 (SLAM: Simultaneous Localization and Mapping) や環境認識などにも多大なニーズがある。そのため、Computer Vision の主要タスクとして数多くの研究が行われており、非常に活発な研究領域である。

古くから、単眼画像から形状を推定する手法として、輝度やテクスチャ、影などの手掛かりを利用する *Shape from X* ( $X \in \{\text{shading, Silhouette, Texture, Focus, etc.}\}$ ) の研究が行われてきた。しかし、より頑健な復元を行うためには、視点移動を伴う複数枚の画像、あるいは動画 (Image sequence) を用いて幾何的な整合性を元に手法が主流となった。この手法、タスクの総称を Structure from Motion (SfM) と呼ぶ。この SfM を起点とする、3D Vision for Geometry 手法の発達、高度な UI を備えた商用ソフト (Pix4D, Metashape, etc.) の出現、オープンソース化により、今日では Photogrammetry 技術は広く一般に普及した。

### 3.1.1 Structure from Motion

項目のタイトルが英語と日本語が混在。日本語がある場合はなるべくした方がよいか

Structure from Motion (SfM) は、複数の視点から撮影された画像群に基づき、カメラの内部パラメータ (Intrinsic Parameters)・外部パラメータ (Extrinsic Parameters、すなわちカメラの三次元的位置と姿勢) と、シーンの疎な (Sparse) 3次元構造を同時に推定する手法である。SfM の理論的基礎は、Tomasi-Kanade による行列分解法 [42] に始まり、Hartley と Zisserman によって多視点幾何学 (Multiple View Geometry) の理論体系 [43] として確立された。2016年に発表された COLMAP [7] は、高い精度

と汎用性、そしてオープンソースプロジェクトとしての完成度から、現在でもアカデミック分野におけるデファクトスタンダードとして広く利用されている。COLMAPにより出力される正確なカメラ内部パラメータと外部パラメータ、および歪み補正済み画像（Undistorted Images）は、後述する密な三次元再構成（Dense Reconstruction）や新規視点合成（Novel View Synthesis）タスクのための入力データとして、重要なパイプラインの一角を担っている。

一般的なインクリメンタル SfM パイプラインでは、まず SIFT などの特徴量検出器および記述子 [44] を用いて、画像間でキーポイント（特徴点）の検出およびマッチングを行う。その後、2 画像間の基本行列  $F$  を、8 点アルゴリズム [45] および RANSAC [46] を組み合わせて推定し、特異値分解（SVD）によるカメラ姿勢の復元へと繋げる。新たなカメラ姿勢は、Perspective-n-Point (PnP) アルゴリズムにより、既に推定された 3 次元点と新規画像内の 2 次元特徴点との対応付きで逐次的に登録する。対応点情報から三角測量（Triangulation）によって追加の 3 次元点群を求め、最後にバンドル調整（Bundle Adjustment, BA）[47] によって再投影誤差を最小化することで、カメラパラメータおよび 3 次元点群の両者を同時に高精度に最適化する。

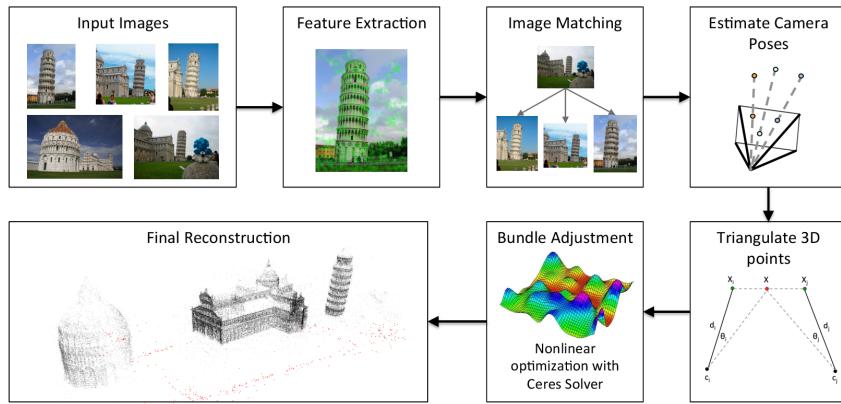


図 3-1: SfM Pipeline [48]

### 3.1.2 Multi-View Stereo

無駄に力入れて書きすぎ。MVS とかほとんど本題じゃない。むしろ SfM を頑張って記述するべきだった。(いい勉強になったが)、Appendix に回した方がいいわ

SfM などによって得られたカメラパラメータが既知の多視点画像群から、各画素単位で密な深度推定により、高密度な三次元復元を行うのが Multi-View Stereo (MVS) である。SfM の出力三次元情報が疎な (Sparse) 点群であるのに対して、MVS は物体表面の密な (Dense) な三次元点群やメッシュ (Mesh) を推定することから、MVS は Dense 3D Reconstruction (密な三次元再構成) のタスクを行っていると言える。これを達成するため、各画素における深度 (Depth) や法線 (Normal) といった幾何パラメータを、画像間の整合性 (Photometric Consistency) に基づいて最適化する手法がとられる。本節では、パッチベース手法の古典である PMVS [8] の概念を概観した後、現在のデファクトスタンダードなパイプライン (COLMAP [49]) の基礎となっている PatchMatch Stereo について詳述する。

## Patch-based Multi-View Stereo (PMVS)



図 3-2: PMVS Pipeline。[48] から引用。左から右にかけてそれぞれの画像は、(1) 入力された多視点画像の一例; (2) 検出された特徴点; (3) 初期マッチング後のパッチ; (4) 拡張とフィルタリング後のパッチ; (5) メッシュモデル。

[Furukawa2010\_PMVS] によって提案された Patch-based Multi-View Stereo (PMVS) は、物体表面を微小な平面パッチの集合としてモデル化し、これを拡張・最適化することで密な 3 次元形状を復元する手法である。本手法は、特徴点マッチングによる疎な初期復元から出発し、信頼度の高いパッチを周囲に拡張(Expand)していくアプローチをとる、物体表面の形状を密に 3 次元推定するための先駆的かつ代表的な手法である。

cite では、現在の設定では名前が表示されずただ、[8] のように番号のみになるため、別の方や引用のフォーマットの設定を考える

PMVSにおける基本単位であるパッチ  $B$  は、単なる画像の矩形領域ではなく、対象物体の表面に接する微小平面（局所接平面）の近似として定義される。パッチ  $B$  は以下のパラメータを持つ。

- 中心座標  $p_B \in \mathbb{R}^3$ : パッチの中心位置。
- 法線ベクトル  $n_B \in \mathbb{R}^3$ : パッチの向きを表す単位法線ベクトル。
- 参照画像  $R_B$ : パッチ  $B$  を観測する画像の中で、光学的・幾何的に最も適した画像。
- 可視画像集合  $V_B$ : パッチ  $B$  がオクルージョンなく観測可能であり、かつ相関スコアが閾値以上となる画像の集合。

PMVS の処理は、特徴点からの初期化(Initialization)の後、以下の 3 つのステップ、拡張(Expansion)、フィルタリング(Filtering)、最適化(Optimization)を反復することで行われる。

1. **Initialization (初期化):** SfM と同様に、各画像から特徴点検出・マッチングを行う [44]。ここから得られる疎な点群を Seed Patch として、初期の法線とともに最初のパッチ群を生成する。前処理としてカメラパラメータの取得に SfM を用いている場合、これらの特徴点と、三角測量(Triangulation)によって得られる疎な三次元点群はそのまま使用できる。
2. **Expansion (拡張):** 物体の表面が滑らかに連続しているという仮定に基づき、パッチをシーンの表面に沿うように増殖させていく。既存のパッチ  $B$  を参照画像  $R_B$  および可視画像集合  $V_B$  に投影

し、その隣接画素に対応する空間領域にパッチが存在しない場合、新たなパッチ  $B'$  を生成する。この際、親パッチ  $B$  の法線  $\mathbf{n}_B$  と深さ情報を初期値として継承させることで、テクスチャが弱い領域であっても、隣接する確度の高い領域から表面を「張り出して」いくことが可能となる。

3. **Filtering (フィルタリング)**: 拡張プロセスによって生じた誤ったパッチを除去する。以下の 3 つの基準が主に用いられる。

- **Visibility Consistency**: 複数のパッチが同一の視線上に存在する場合、カメラに近い方を残し、隠蔽される奥のパッチを削除する。
- **Photometric Consistency**: 正規化相互相關 (NCC) 等を用いた画像間の整合性スコアが一定以下のパッチを外れ値として破棄する。
- **Number of Views**: パッチを安定して観測できるカメラの台数  $|V(p)|$  が最小閾値未満のものを信頼性不足として削除する。

4. **Optimization (最適化)**: 各パッチの位置  $p_B$  と法線  $\mathbf{n}_B$  を微修正し、画像間の整合性を最大化する。具体的には、パッチ  $B$  を可視画像  $I \in V_B$  へ投影して得られる画素値と、参照画像  $R_B$  上の画素値との間の Photometric Discrepancy を最小化するように、非線形最適化を行う。

$$g(p) = \frac{1}{|V(B) \setminus R(B)|} \sum_{I \in V(B), I \neq R(B)} g(B, I, R(B))$$

ここで、 $g(B, I_1, I_2)$  は、パッチ  $B$  に対する、画像  $I_1$  と  $I_2$  の間のフォトメトリック不一致度 (Photometric Discrepancy) を測定するための関数であり、正規化相互相關 (NCC) などを使用することができます。フォトメトリック不一致度は、同一の 3 次元パッチを異なる視点から観測した際の画素値の差異を定量化する指標である。この際、パッチの法線  $\mathbf{n}(p)$  を考慮して各画像をホモグラフィ変換することで、視点による透視歪みを補正し、より正確なマッチングを実現している。

以上のプロセスを収束するまで繰り返すことで、初期の疎な点群は徐々に密度を増し、最終的に物体表面全体を覆う密な点群が得られる。**[Furukawa2010\_PMVS]** の手法は、大域的な最適化を行う PatchMatch Stereo 等と比較して局所的な貪欲法に近い性質を持ち、計算コストが高いが、拡張ステップによる表面の連続性利用が強力であり、密な復元を実現するための歴史的に重要な手法である。

### 3.1.3 PatchMatch Stereo

PMVS が、信頼できる「種 (Seed)」から局所的に表面を拡張していくアプローチであるのに対し、**[50]** が提案した PatchMatch Stereo は、画像上のすべての画素に対して、個別の 3 次元平面パラメータ (深度と法線) を推定する手法である。

従来の局所ステレオマッチング手法は、マッチングウィンドウ内の深度が一定である（カメラに対して平行な平面：Fronto-parallel window）と仮定することが一般的であった。しかし、この仮定は傾いた面や曲面において成立せず、再構成精度の低下や「階段状」のアーティファクトを生む原因となっていた。

これに対し、Bleyer らは各画素のウィンドウを 3 次元空間内の傾いた平面（Slanted Support Window）としてモデル化した。しかし、各画素に対して最適な平面パラメータ（深度および法線の向き）を決定しようとすると、その探索空間は連続値であり無限大となるため、従来の総当たり的な探索や離散ラベルを用いる手法（Graph Cuts など）は適用できない。

この問題を解決するために導入されたのが、Barnes らによる最近傍探索アルゴリズム ”PatchMatch” の概念をステレオ視に応用した推論フレームワークである。

**平面モデルとマッチングコスト:** 左画像の各画素  $\Gamma$  に対し、3 次元平面  $f_\Gamma$  を割り当てる。平面  $f_\Gamma$  は 3 つのパラメータ  $(a_{f_\Gamma}, b_{f_\Gamma}, c_{f_\Gamma})$  を持ち、画素  $\Gamma$  の座標  $(\Gamma_x, \Gamma_y)$  における深度  $d_\Gamma$  は以下の式で表される。

$$d_\Gamma = a_{f_\Gamma} \Gamma_x + b_{f_\Gamma} \Gamma_y + c_{f_\Gamma}$$

このモデルにより、画素ごとに異なる法線を持つ傾いた平面を表現でき、サブピクセル精度の深度推定が可能となる。最適化の目的は、各画素  $\Gamma$  において、アグリゲーションコスト  $m(\Gamma, f_\Gamma)$  を最小化する平面  $f_\Gamma$  を、無限の候補空間  $\mathcal{F}$  から見つけ出すことである。

$$f_\Gamma = \underset{f \in \mathcal{F}}{\operatorname{argmin}} m(\Gamma, f)$$

**ランダム探索と伝播による推論:** PatchMatch Stereo の核心は、ランダムな初期化状態から、空間的・視点的な相関を利用して「良い解」を画像全体に伝播(Propagate)させるプロセスにある。アルゴリズムは、以下のステップを反復することで収束する。

1. **Random Initialization (ランダム初期化):** 初期状態では、すべての画素に対し、ランダムなパラメータを持つ平面（ランダムな深度と法線ベクトル）を割り当てる。
2. **Spatial Propagation (空間伝播):** 「隣接する画素は、同じ平面上に乗っている可能性が高い」という仮定を利用する。画像走査順序に従い、現在の画素  $\Gamma$  とその近傍画素  $\Gamma'$ （例えば左上の画素）を比較する。もし、近傍画素  $\Gamma'$  の持つ平面  $f_{\Gamma'}$  を画素  $\Gamma$  に適用した際のマッチングコストが、現在の平面  $f_\Gamma$  よりも低くなるならば、画素  $\Gamma$  は平面  $f_{\Gamma'}$  を自身の新たな推定値として採用し、コピーする。これにより、画像の一部で偶然「正解に近い平面」が見つかれば、その平面情報は波紋のように隣接画素へと広がり、領域全体が正しい平面で埋め尽くされていく。
3. **View Propagation (視点間伝播):** ステレオ特有の拡張として、左右画像間の一貫性を利用する。左画像の画素  $\Gamma$  の対応点である右画像の画素  $\Gamma'$  が、より適切な平面パラメータを持っている場合、それを左画像座標系へ変換して取り込む。これにより、オクルージョン領域外での整合性が強力に担保される。
4. **Plane Refinement (平面の微調整):** 伝播のみでは、既存の平面パラメータのコピーしか行われないため、真値へ到達できない。そこで、現在の平面パラメータに微小なランダム擾動 (Perturbation) を加え、コストが改善するかをテストする。反復が進むにつれて擾動の範囲を指數関数的に狭めていくことで、サブピクセルレベルでの高精度な収束を実現する。

この手法の利点は、巨大なコストボリューム (Cost Volume) をメモリ上に構築する必要がないため、高解像度画像や大きな視差範囲に対してもメモリ効率が良い点にある。また、連続空間での最適化を行うため、離散化に伴う量子化誤差が発生せず、極めて滑らかな曲面や急峻な傾斜面の復元に成功している。

[50] は、基本的に2枚の画像（ステレオペア）間でのマッチングを前提としている。しかし、実際のSfM/MVSパイプラインでは、数十から数千枚の画像（Multi-View）が入力され、かつそれらがインターネット上の写真のように撮影条件がバラバラな「非構造化（Unstructured）」データである場合もある。この課題に対し、[49] は、PatchMatch Stereo のフレームワークを多視点へ拡張する際、「画素ごとの視点選択（Pixelwise View Selection）」という概念を導入することで解決を図った。

多視点ステレオにおける最大の課題は、ある参照画像の画素  $\Gamma$  を復元するために、「どのソース画像を使うべきか」が画素ごとに異なる点である。画像全体で一律にソース画像を選んでしまうと、オクルージョンや解像度の不一致により、特定の画素ではマッチングが破綻してしまう。COLMAPでは、以下の3つの幾何学的事前分布 (Geometric Priors) を確率モデルに組み込み、各画素  $\Gamma$  が自分自身にとって最適なソース画像を動的に選択しながら推論を行う。

1. **Triangulation Prior:** 十分なベースラインを持ち、三角測量の精度が保証される角度 (Triangulation Angle) で撮影された画像を優先する。角度が小さすぎる（視点が近すぎる）画像は深度推定の不確定性が高いため除外される。
2. **Resolution Prior:** 参照画像とソース画像で、対象を捉えている解像度（画素密度）が類似している画像を優先する。極端に解像度が異なる画像間でのマッチングは、エイリアシング等の問題を引き起こすためである。
3. **Incident Prior:** 推定された法線ベクトルに対し、カメラ視線が正対に近い（斜めすぎない）画像を優先する。これにより、極端な浅い角度から撮影された信頼性の低い画像の影響を排除する。

Schönberger らは、これらの幾何学的尺度とフォトメトリックな整合性を統合した確率的グラフィカルモデルを構築し、PatchMatch の反復プロセスの中で「深度・法線の推定」と「最適なソース画像の選択」を同時に最適化する手法を確立した。これにより、COLMAP は極めてノイズの多い非構造化データセットに対しても、ロバストな密な三次元再構成を可能にした。

一方、これらの手法は幾何光学 (Geometric Optics) による光の直進性を仮定しているが、これは屈折のあるシーンでは成立しない。また、反射 (Reflection) を含むシーンにおいても、従来の手法は一般的に困難を抱えている。

### 3.1.4 Feed Forward 3D Reconstruction

本研究の本筋である最適化ベース (Optimization-based) の手法とは対照的に、近年 (2023-2025年)、大規模データセットと Deep Learning、特に Transformer アーキテクチャの発展により、幾何学的計算を推論 (Inference) として解く「フィードフォワード型 (Feed-Forward)」の手法が急速に台頭している。従来の手法 (SfM/MVS や NeRF/3DGS) が、入力シーンごとにパラメータを反復的に更新して解を探索

するのに対し、フィードフォワード型の手法は、学習済みの膨大な事前知識（Priors）を用いて、単一の順伝播処理のみで3次元構造を回帰するデータ駆動型（Data-driven）のアプローチである。

このアプローチの代表例として、以下の手法が挙げられる。

- **Depth Anything** [yang2024depthanything]: 6200万枚以上の画像から学習された基盤モデル（Foundation Model）であり、DINOv2 バックボーンを活用することで、テクスチャのない領域や未知のシーンに対しても極めてロバストな単眼深度推定を実現している。しかし、出力はスケール不確定性を伴う2.5次元表現に留まり、多視点間での厳密な幾何学的整合性は保証されない。
- **DUSt3R / MASt3R** [wang2024dust3r leroy2025mast3r]: 従来のSfMパイプライン（特徴点抽出、マッチング、バンドル調整）を完全に排除し、2枚の画像から直接「ポイントマップ（Point Map）」を回帰する手法である。これにより、カメラパラメータを事前に与えることなく（Unposed）、エンドツーエンドでの3次元形状およびPoint Mapからのカメラ姿勢の逆推定が可能となった。
- **VGGT (Visual Geometry Grounded Transformer)** [wang2025vggt]: CVPR 2025にてBest Paper Awardを受賞した、現時点での到達点といえる手法である。VGGTは、任意の枚数の画像を入力とし、カメラパラメータ、深度、点群、そして追跡情報（Tracks）の全てを同時に推論する。

これらの手法は、計算速度とロバスト性において革新的であるが、本研究が扱う「物理的に正確な表面再構成」の観点からは明確な限界も存在する。フィードフォワード手法はその性質上、学習データに深く性能を依存する。

そのため、屈折（Refraction）や透明物体（Transparency）を含むシーンにおいて、これらの手法は破綻しやすい。一般的な学習データセット（ScanNet等）には透明物体が十分に、かつ物理的に正確なアノテーションと共に含まれていないため、フィードフォワードモデルはガラス表面を背景と混同したり、深度を平滑化してしまう傾向がある。対して、本研究で用いるGaussian Splatting等の最適化ベース手法は、屈折率（Index of Refraction: IoR）やスネルの法則を明示的にモデル化することで、こうした物理現象を正確に逆算することが可能である。

しかし、動画生成AIが流体や光の反射といった物理法則をデータから獲得しつつある現状[quan2025transparent]を鑑みると、将来的にはフィードフォワード手法も十分なデータスケールによって屈折を「学習」する可能性は否定できない。現時点では、幾何学的整合性と物理的忠実性を担保するためには、依然として物理モデルに基づく最適化が不可欠である。

### 3.2 Inverse Rendering as 3D Reconstruction

#### Inverse Rendering

<https://gemini.google.com/app/0bb4c009393122eb>

コンピュータグラフィックス(CG)の古典的な課題は、3次元のシーン記述(形状、材質、光源)から2次元の画像を生成することである。この過程は「Forward Rendering(順方向レンダリング)」と呼ばれ、一般的にはレンダリングと呼ぶ。これを数学的な関数  $R$  と見なすと、シーンパラメータ  $\Theta$  から画

像  $I$  への写像  $I = R(\Theta)$  として表現できる。これに対し、Inverse Rendering（逆レンダリング）は、観測された画像  $I_{obs}$  から、それを生成したシーンパラメータ  $\Theta$  を推定する逆問題  $\Theta = R^{-1}(I_{obs})$  を解くことである。観測された画像から 3D シーンを推定するという意味で、Inverse Rendering は画像からの三次元再構成の一例であるといえる。

しかし、物理的な世界から画像への射影は情報の損失（奥行きの消失、法線とテクスチャの混同など）を伴うため、この逆問題は本質的に不良設定問題（Ill-posed problem）となる。この困難な問題を、最適化手法を用いて解くための強力なフレームワークとして登場したのが、Differentiable Rendering（微分可能レンダリング）である。



### 3.2.1 Analysis-by-Synthesis and Novel View Synthesis

現代の Inverse Rendering の多くは、Analysis-by-Synthesis というアプローチを採用している。これは、パラメータ  $\Theta$  を直接回帰するのではなく、レンダリングされた画像  $R(\Theta)$  と観測画像  $I_{obs}$  との間の再構成誤差  $\mathcal{L}$  を最小化する最適化問題として定式化される。

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} \mathcal{L}(R(\Theta), I_{obs})$$

ここで  $\mathcal{L}$  は損失関数 (L2 ノルムや Perceptual Loss) である。この最小化問題を勾配降下法 (Gradient Descent) で解くためには、レンダリング関数  $R$  がパラメータ  $\Theta$  に関して微分可能である必要がある。すなわち、3D シーンパラメータ  $\Theta$  に関する損失関数  $\mathcal{L}$  の勾配  $\frac{\partial \mathcal{L}}{\partial \Theta}$  を計算できなければならない。連鎖律 (Chain Rule) を適用すると、勾配は以下のように分解される。

$$\frac{\partial \mathcal{L}}{\partial \Theta} = \frac{\partial \mathcal{L}}{\partial I} \cdot \frac{\partial I}{\partial \Theta}$$

前半の  $\frac{\partial \mathcal{L}}{\partial I}$  (画像の画素値に対する Loss の勾配) は容易に計算できるが、後半の  $\frac{\partial I}{\partial \Theta}$  (シーンパラメータの変化が画素値にどう影響するか) の計算は一般的な三次元形状表現である Mesh では難しい。  
いらないと思う

第 3.2.1 小節の定式化によって、

#### 微分可能レンダリングによる最適化

逆レンダリングを勾配法 (Gradient Descent) によって解くためには、レンダリングプロセスが微分可能 (Differentiable) である必要がある。(なんで勾配法で解く必要があるのか土木人に分かるよう

に説明してあげないと。DL の発達とその最適化手法の根底にあることを言及) 微分可能レンダリング (Differentiable Rendering, DR) の枠組みでは、以下のループによって 3 次元シーンが最適化される。

1. **レンダリング:** 現在推定されている 3 次元シーン (形状・外観) から、仮想的な視点で画像を生成する。
2. **損失計算:** 生成された画像 (Rendered Image) と、実際に撮影された正解画像 (Source Image) との間の誤差 (Loss) を計算する。
3. **誤差逆伝播:** 微分可能なレンダラを介して、誤差の勾配 (Gradient) を 3 次元シーンのパラメータへと連鎖律に従って伝播させる。
4. **更新:** 勾配を用いて、3 次元シーンのパラメータを逐次更新する。

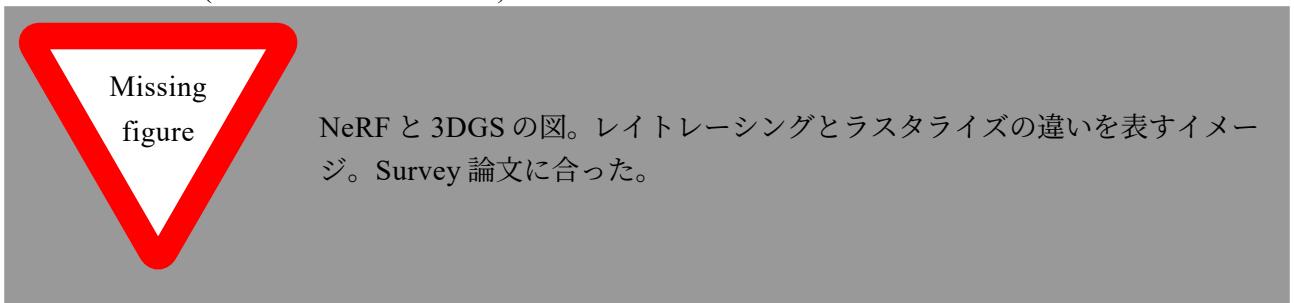
この手法は、幾何学的な特徴点のみならず、画像内の全画素の情報を最適化に利用できる従来の Bundle Adjustment も 2 次元画像座標上の特徴点を介して、3 次元シーンを構成する点群を再投影誤差を介し最小化するという点で、微分可能レンダリングの一種と捉えることができる。

### Novel View Synthesis

従来の SfM や MVS の手法は Geometric Reconstruction として、主に幾何情報の復元に焦点を当てていた。実際の Scene は幾何情報 (Geometry) に加え、照明 (illumination)、テクスチャ (texture)、BRDF などでモデル化される表面特性などを含む。新規視点合成というタスクでは、ポーズを付与した画像から、その視点外からの新視点の画像を推定するタスクである。SfM や MVS が Geometric Reconstruction とするならば、新規視点合成は Appearance Reconstruction となる。Appearance には当然、Geometry の情報も含まれているため、Appearance Reconstruction は Geometry Reconstruction のタスクを暗に含んでいると言える。図を挿入して上げるぞ。(CG のレンダリングプロセスと、ベン図)

### Neural Radiance Fields (NeRF)

<https://dl.acm.org/doi/10.1145/3758085> [51] からまとめる。微分可能レンダリングの代表例が NeRF (Neural Radiance Fields) である。



NeRF では、シーンをボクセルやポリゴンといった明示的な幾何構造ではなく、多層ペーセプトロン (MLP) を用いた「連続的な輝度場 (Radiance Field)」として表現する。これは、点群やメッシュといっ

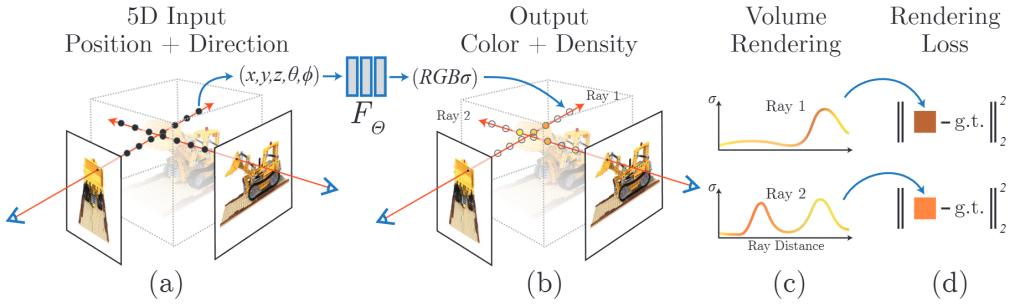


図 3-3: NeRF によるシーン表現と微分可能レンダリングの処理の概念図。[19]

から引用。(a) カメラ光線に沿って、5 次元座標(位置および視線方向)をサンプリング。(b) それらを MLP に入力し、色と volume density を出力させます。(c) ボリュームレンダリングを用い、画像を生成する。(d) 合成された画像と実際に観測された GT 画像との間の誤差を最小化することによって、シーン表現の最適化を行う。

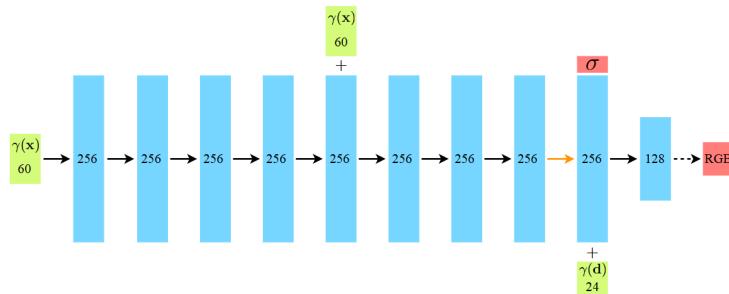


図 3-4: NeRF のネットワークアーキテクチャ。[19]

から引用。ネットワークアーキテクチャの詳細。入力となる 5 次元座標のうち、位置  $\mathbf{x}$  は 8 層の全結合層によって処理され、体積密度  $\sigma$  と 256 次元の特徴ベクトルを出力する。この特徴ベクトルは視線方向  $\mathbf{d}$  と結合され、続く 1 層の全結合層を経て、視点依存 (view-dependent) の RGB 色が出力される。全ての層において活性化関数として ReLU を使用する。

た、シーンの幾何情報を直接的にエンコードする Explicit Representation ではなく、関数を介して表現する Implicit Representation の一種である。

Missing figure

NeRF の MLP の構造も図で載せたい。引用できる

具体的には、空間上の座標  $(x, y, z)$  と視線方向  $(\theta, \phi)$  を入力とし、その点における放射輝度 (RGB) と体積密度 ( $\sigma$ ) を出力する関数  $f_\Theta$  を学習する。この表現を用いて、以下の Volumetric Rendering によりピクセル値を算出する。

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \sigma(\mathbf{r}(t)) \mathbf{c}(\mathbf{r}(t), \mathbf{d}) dt$$

ここで  $T(t)$  は透過率、 $\mathbf{r}(t)$  はレイ上の点を表す。この積分プロセスは離散的なサンプリングによっ

て近似され、微分可能な形で実装される。ボリュメトリックレンダリングは、元来、煙や炎などの不均等な媒質を通過する光の伝播をモデル化を可視化するために開発されたものだが、メッシュなどのラスター化に含まれる離散性がなく、微分可能性を保つとともに、汎用的な Scene 表現に用いられる。これにより、複雑な光学的特性（反射や半透明など）を含むシーンにおいても、エンドツーエンドでの高精度な再構成が可能となった。

ピクセル色の決定には、古典的なボリュームレンダリングの原理が応用される。仮想的なカメラから放たれたレイ（光線）に沿って空間をサンプリングし、各サンプル点での密度  $\sigma$  を重みとした放射輝度  $c$  の積分値を算出することで、最終的なピクセル値が決定される。この積分計算は離散的な総和として近似されるが、演算過程の全てが微分可能に保たれているため、レンダリング画像と実画像の二乗誤差を損失関数とし、MLP の重みをエンドツーエンドで最適化できる。しかし、NeRF は写真のような忠実度の自由視点合成を実現した一方で、実用上の重大な障壁も露呈させた。1 ピクセルの描画のためにレイに沿った数百回の MLP 推論を要する計算負荷の高さは、リアルタイムレンダリングを困難にし、またシーンごとに日単位の学習を要する点は、大規模なデータセットへの適用を制限している。また、3 次元シーンが MLP にエンコードされる点で、その解釈や編集可能性が乏しく、測量用途としての適用性に欠ける。これらの課題、すなわち「計算資源の集約性」と「暗黙的表現による編集の困難性」を克服しようとする動機が、後の明示的な幾何プリミティブを用いる手法への回帰を促すこととなった。

### 3.3 Gaussian Splatting

#### 3D Gaussian Splatting

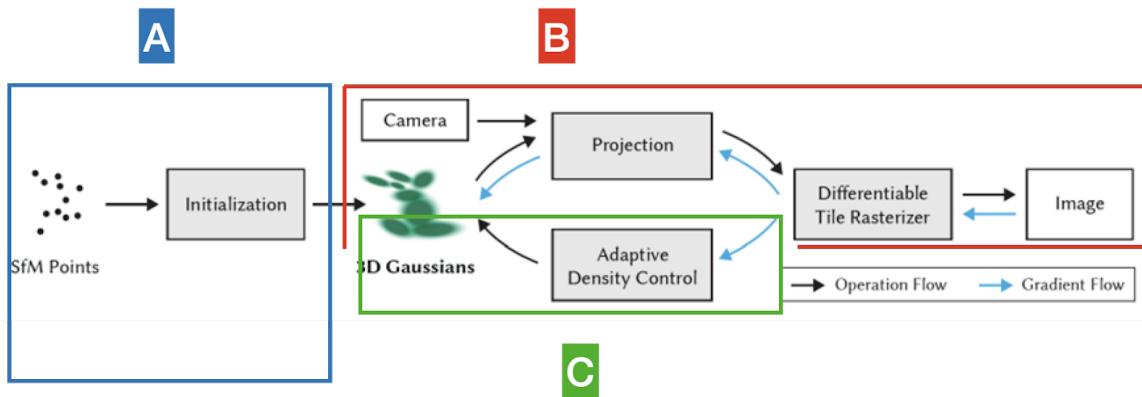


図 3-5: Gaussian Splatting のパイプライン。[20] より引用。

Missing figure

NeRF と 3DGS の分類図。レンダリング速度と、データサイズの 2D プロット。Barron さんのレクチャー動画から。学習速度も入れれば、2 次元に収まらない。。2 個いるかな？ できれば、MVS とかも含めて示せねばいい。それだったら表が美しい

3D Gaussian Splatting (3DGS) は、新規視点合成において最先端 (SOTA) の結果を達成している近

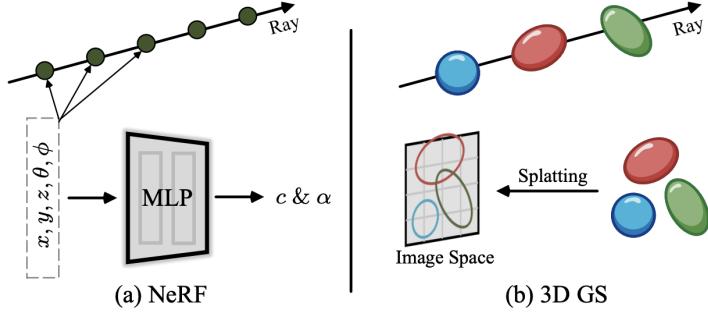


図 3-6: Gaussian Splatting と NeRF とレンダリング方式の違い。[52] より引用。(a) NeRF のレンダリング方式。(b) 3DGS のレンダリング方式。NeRF では各レイに沿ってサンプリングするため、計算コストが大きいが、GS ではラスタライズによる高速なレンダリングを実現する。これは GPU の描画パイプラインを効果的に流用する。ラスタライズにより 3D 空間の Gaussian を 2D 画像平面に投影することを ‘Splatting’ と表現している。

年の手法である [20]。その特徴は、フォトリアリスティックかつ高忠実度な 3 次元シーンのキャプチャ能力、高速な学習時間、そしてリアルタイムレンダリングにある。明示的な(Explicit)3 次元表現である 3DGS は、Visual-SLAM [53–55]、アバター生成 [56·57]、フィードフォワード型 3 次元再構成 [58] など、幅広いタスクへの適用に成功している。その可能性はさらに広がり、衛星画像からの数値表層モデル(DSM)生成 [59]、自動運転、そして水中 3 次元再構成 [60] といった様々な実世界アプリケーションへと拡張されている。

### 自動運転



図 3-7: 点群と樁円体のレンダリング結果の比較。(a) 点群のレンダリング結果。(b) 樁円体のレンダリング結果。(c) 最終的な 3DGS のアルファブレンディングによるレンダリング結果。

3DGS のパイプラインは主に、レンダリングを行うフォワードパスと、最適化を行うバックワードパスの 2 つの段階で構成される。フォワードパスでは、3 次元ガウス分布(3D Gaussians)の集合を、ラスタライズにより画像平面に投影することで画像を合成する。各ガウス分布は、中心位置  $p \in \mathbb{R}^3$ 、不透明度  $\alpha \in [0, 1]$ 、球面調和関数(SH)によって表現される視点依存の色係数  $c(p, t_i) \in \mathbb{R}^3$ 、および 3 次元共分散行列  $\Sigma^{3D} \in \mathbb{R}^{3 \times 3}$  という、最適化可能なパラメータセットによって定義される。これらのパラメータは、

カメラパラメータを得る際に用いられる SfM で副次的に得られる粗点群を用い初期化される(図 3-5A)。共分散行列  $\Sigma^{3D}$  は、スケーリングベクトル  $s \in \mathbb{R}^3$  から構成されるスケーリング対角行列  $S \in \mathbb{R}^{3 \times 3}$  と、回転クォータニオン(回転行列  $R \in SO(3)$  として表現)を用いて以下のように構成される:

$$\Sigma^{3D} = RSS^\top R^\top$$

点  $x \in \mathbb{R}^3$  に対する対応する非正規化ガウス分布関数は以下で与えられる:

$$G(x) = \exp\left(-\frac{1}{2}(x - p)^T (\Sigma^{3D})^{-1} (x - p)\right)$$

フォワードパスにおいて、あるカメラ視点からの画像をレンダリングするために、これらのガウス分布はまず外部パラメータ行列  $[W|t]$  を用いてワールド座標系からカメラ座標系へと変換される。ガウス分布の中心位置  $p$  と 3 次元共分散行列は以下のように更新される:

$$\begin{aligned} p_{\text{cam}} &= Wp + t \\ \Sigma_{\text{cam}}^{3D} &= W\Sigma^{3D}W^\top \end{aligned}$$

ここで、 $W_{\text{view}} \in \mathbb{R}^{3 \times 3}$  は視点回転行列、 $t \in \mathbb{R}^3$  は平行移動ベクトルである。[\[61\]](#) の提案した射影手法に従い、カメラ空間における 3 次元共分散行列  $\Sigma_{\text{cam}}^{3D}$  は 2 次元画像平面へと射影される。これは透視投影の一次近似(アフィン近似)のヤコビ行列  $J$  を用いて行われ、2 次元共分散行列  $\Sigma^{2D}$  が得られる:

$$\Sigma^{2D} = J\Sigma_{\text{cam}}^{3D}J^\top$$

各ピクセルの最終的な RGB 値  $\Gamma \in \mathbb{R}^3$  は、射影されたガウス分布をアルファブレンディングすることでレンダリングされる。ピクセルと重なるガウス分布の集合は、まず深度に基づいて手前から奥へとソートされ、視点依存色が以下のように累積される:

$$\begin{aligned} \Gamma(x) &= \sum_{k=1}^K c_k \alpha_k^{\text{pixel}} \prod_{j=1}^{k-1} (1 - \alpha_j^{\text{pixel}}) \\ \text{where } \alpha_k^{\text{pixel}} &= \alpha_k G_k^{2D} \end{aligned}$$

ここで、 $k$  はピクセルに重なる整列されたガウス分布の集合のインデックスである。

バックワードパスでは、最適化により測光誤差(Photometric loss)を最小化する。これは  $\mathcal{L}_1(\Gamma, \Gamma_{gt})$  損失と D-SSIM 損失  $\mathcal{L}_{\text{D-SSIM}}(\Gamma, \Gamma_{gt})$  [\[62\]](#) の加重和である:

$$\mathcal{L} = (1 - \lambda)\mathcal{L}_1 + \lambda\mathcal{L}_{\text{D-SSIM}}$$

したがって、最適化問題は以下のように定式化される:

$$\underset{p, R, s, c, \alpha}{\operatorname{argmin}} \quad \mathcal{L} = \mathcal{L}(\Gamma, \Gamma_{gt})$$

これらの定式化により、パイプライン全体が完全微分可能となり、パラメータ  $\Theta = \{\mathbf{p}, \mathbf{R}, \mathbf{s}, \mathbf{c}, \alpha\}$  は Adam を用いた勾配降下法 [63] によって最適化可能となる。(図 3-5B)

これらの勾配降下法のみでは、特に観測視点が少ない箇所 (Few Shot Area) などで、シーンの適切な一貫性に欠ける局所最適解に陥りやすい。これを解決するために 3DGS では Adaptive Density Control(ADC) というヒューリスティックな手法が導入されている。ある Gaussian に関して、その 2D 平面や 3D 空間における位置に対する誤差勾配  $\frac{\partial \mathcal{L}}{\partial p}$  が大きい場合、その Gaussian は幾何的な誤差の大きいものであるという推察に基づいて分割 (Split)、複製 (Clone)、消去 (Prune) のいずれかの操作を行う。位置誤差勾配がある閾値を超え、さらにシーンに対して Gaussian が大きい場合は分割、小さい場合は複製を行う。また、不透明度  $\alpha$  が小さい Gaussian も消去する。効果的にシーンに寄与しない Gaussian を削除するため、一定のイテレーションごとにシーンの Gaussian の不透明度を一様に小さいものにリセットすることで、その後不透明度が向上しない Gaussian を削除できるようにする。

MCMC の説明も。?。これらのヒューリスティックスはデータセットや初期値に対して依存するため、それを克服するような MCMC といった手法も考案されている。

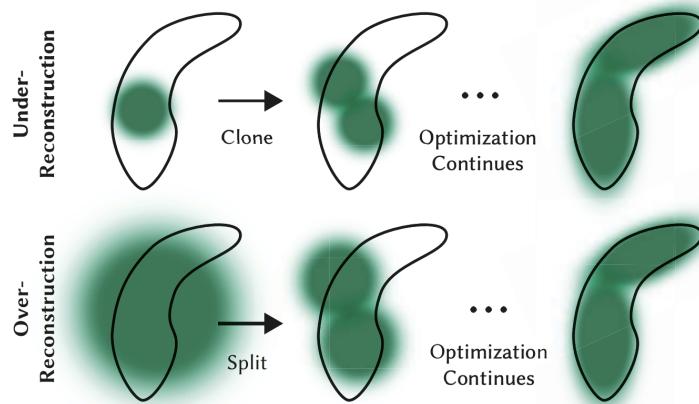


図 3-8: Adaptive Density Control(ADC) の操作。[20] より引用。Gaussian に流れる位置誤差勾配が大きい場合、サイズが小さい Gaussian に対しては、複製、サイズが大きい Gaussian に対しては、分割を行うことで、適切なサイズの Gaussian でシーンを満たす。

---

**Algorithm 1:** Novel 3DGS[20] における密度最適化アルゴリズム。[64] より引用。

---

**Data:** Scene of Gaussians  $G$

```
for  $G_k \in G$  do
    // Densification
    if  $\tau_k \geq T_{grad}$  then
        if  $\max(s_k) > P_{dense} \cdot e_{scene}$  then
            | splitGaussian( $G_k$ )
        end
    else
        | cloneGaussian( $G_k$ )
    end
end
// Opacity Pruning
if  $\alpha_k < \alpha_{min}$  then
    | pruneGaussian( $G_k$ )
end
// Size Pruning
if  $\max(s_k) > 0.1 \cdot e_{scene}$  then
    | pruneGaussian( $G_k$ )
end
end
```

---

ここで、 $\tau_k$  は 2 次元位置誤差勾配の平均値、 $P_{dense}$  は密度最適化のハイパーパラメータ、 $T_{grad}$  は位置誤差勾配の閾値ハイパーパラメータ、 $e_{scene}$  はシーンの範囲、 $\alpha_{min}$  は最小不透明度ハイパーパラメータである。

その最適化に要する学習時間は、30k のイテレーションによって 1 時間以内となり、当時 NVS の Sota であった Mip-NeRF360 [65] に比較して 10 倍以上の高速化を達成した。加えて、Ray-Tracing に比較し、既存の GPU 描画パイプラインの性能を引き出す Tile-Based レンダリングによって、100 fps 以上のリアルタイムレンダリングを実現したことで、インタラクティブな Scene の可視化が可能となった図 3-6。このプロセスを通じて得られた 3 次元ガウス分布の集合は、3 次元シーンを高忠実度で捉えることができる。

しかし、このパイプライン全体はピンホールカメラモデルと透視投影に依存しており、光が直線的に進むことを根本的な前提としている。この前提は、空気と水の境界での屈折が幾何学的矛盾を引き起こすような複数の媒質が介在する環境においては成立しない。この制限にもかかわらず、3DGS の明示的な 3 次元表現は、屈折の法則を数学的に定式化し、シーン表現の幾何学的パラメータそのものに直接適用することが可能とする。一方で、3DGS ではラスタライズによってレンダリングされるため、レイトレイシングのように容易に屈折の影響をモデル化することができない、という困難も生じる。

### 3.3.1 2D Gaussian Splatting

2D Gaussian Splatting(2DGS) [66] は、Photometric な一貫性に特化した 3DGS に対し、幾何的な一貫性を向上させることを目的とした派生手法である。3 次元シーンで観測される Radiance は、環境光が物体表面で反射したものから観測されるという前提から、3 次元シーンの各 Gaussian は物体表面に整列して配置されることを目指す MVS などの Geometric Reconstruction の一種とも言える。

2DGS の 3DGS との違いは、主に以下の 3 点である：

- 平らな 2 次元ガウス分布 (Surfel) プリミティブによる 3 次元シーン表現
- 正確な透視投影 (Perspective-accurate) による、2D Gaussian 集合の 2 次元画像平面への射影
- 幾何的な整合性を高める正則化の導入

2DGSにおいて、シーンは平らな 2 次元ガウス分布 (Surfel)(2D-oriented planar Gaussian disks) の集合として表現される。つまり、3DGS に比較し、スケーリングベクトル  $s$  のうち最小の成分が 0 と表現した Surfel として表現し、2DGS のプリミティブは、中心位置  $\mathbf{p}_k$ 、主方向ベクトル  $\mathbf{t}_u$  と  $\mathbf{t}_v$ 、および対応するスケーリングベクトル  $s = (s_u, s_v)$  によって表現される。(回転行列  $\mathbf{R}_k$ 、不透明度  $\alpha_k$  と視線依存色  $c_k$  は 3DGS と同様である。)

具体的には、3D 空間内の 2 次元ガウス分布は、ローカルな接平面  $(u, v)$  上で定義され、その幾何学的形状は同次変換行列  $\mathbf{H} \in \mathbb{R}^{4 \times 4}$  を用いて以下のようにパラメータ化される [66]。

$$\mathbf{H} = \begin{bmatrix} s_u \mathbf{t}_u & s_v \mathbf{t}_v & \mathbf{0} & \mathbf{p}_k \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

ここで、ローカル座標  $\mathbf{u} = (u, v)$  におけるガウス分布の強度は、標準的な 2 次元ガウス関数として定義される。

$$G(\mathbf{u}) = \exp\left(-\frac{u^2 + v^2}{2}\right)$$

### Perspective-Accurate Splatting via Ray-Splat Intersection

3DGS が採用しているアフィン近似による投影は、ガウス分布の中心から離れるにつれて近似誤差が増大し、視点によって深度が不整合になるという問題がある。これに対し 2DGS は、視線レイ (Ray) と 2 次元平面 (Splat) の交差判定 (Ray-Splat Intersection) を明示的に計算することで、正しい透視投影を実現する。

スクリーン空間上のピクセル座標  $\mathbf{x} = (x, y)$  を通る視線レイは、3 次元空間において直交する 2 つの平面 (x-plane と y-plane) の交線として定義できる。これを同次座標系の 4 次元平面ベクトル  $\mathbf{h}_x = (-1, 0, 0, x)^\top, \mathbf{h}_y = (0, -1, 0, y)^\top$  として表現し、これらをワールド空間からスクリーン空間への変換行列  $\mathbf{W}$  と、先述のローカル変換行列  $\mathbf{H}$  を用いてガウス分布のローカル座標系へ逆変換する。ここで、平面の変換には逆転置行列  $(\mathbf{WH})^{-\top}$  を用いる必要があるが、これにより明示的な逆行列計算を回避し、数値的な安定性を保つことができる。

$$\mathbf{h}_u = (\mathbf{W}\mathbf{H})^\top \mathbf{h}_x, \quad \mathbf{h}_v = (\mathbf{W}\mathbf{H})^\top \mathbf{h}_y$$

この変換された平面  $\mathbf{h}_u, \mathbf{h}_v$  と、ローカル平面上の点  $(u, v, 1, 1)^\top$  との内積が 0 になるという条件から、交差点  $\mathbf{u}(\mathbf{x})$  を以下の閉形式で高速に求めることができる。

$$\mathbf{u}(\mathbf{x}) = \frac{1}{\mathbf{h}_u^3 \mathbf{h}_v^4 - \mathbf{h}_u^4 \mathbf{h}_v^3} \begin{pmatrix} \mathbf{h}_u^2 \mathbf{h}_v^4 - \mathbf{h}_u^4 \mathbf{h}_v^2 \\ \mathbf{h}_u^4 \mathbf{h}_v^1 - \mathbf{h}_u^1 \mathbf{h}_v^4 \end{pmatrix}$$

ここで  $\mathbf{h}^i$  はベクトルの第  $i$  成分を表す。これにより得られた  $\mathbf{u}(\mathbf{x})$  を用いて  $G(\mathbf{u}(\mathbf{x}))$  を評価することで、正確なレンダリングが可能となる。

また、2DGS 特有の問題として、視線が円盤に対して平行に近い角度 (Slanted viewpoint) になった際に、スクリーン上で線分に退化し、勾配消失やエイリアシングを引き起こす現象がある。これを防ぐため、2DGS ではローパスフィルタ  $\sigma$  を導入し、スクリーン空間での最小サイズを保証する以下のクリッピング処理を行う。

$$\hat{G}(\mathbf{x}) = \max \left\{ G(\mathbf{u}(\mathbf{x})), G\left(\frac{\mathbf{x} - \mathbf{c}}{\sigma}\right) \right\}$$

この処理は実装上、最適化の安定性を確保するために極めて重要である。

ローパスフィルタまでは説明しなくても良い。(3DGS でも説明していないから)。MipSplatting 説明しないといけない

## Regularization for Geometric Consistency

2DGS は Surfel 表現を持つが、単なる Photometric loss のみでは、3 次元再構成特有の不定性によりノイズの多い表面形状に収束してしまう傾向がある。そこで、幾何学的精度を向上させるために [65] で提案された 2 つ正則化項を GS フレームワークに導入している。

一つ目は **Depth Distortion Loss** ( $\mathcal{L}_d$ ) である。3DGS のボリュームレンダリングは、レイ上の交点間の距離を考慮しないため、ガウス分布が前後に散らばるアーティファクトが生じやすい。2DGS では、レイ上の交差点深度  $z_i$  とブレンディング重み  $\omega_i$  を用いて、寄与の高いガウス分布同士の深度距離を最小化するように制約をかける。

Missing figure

依然、木津川で再構成した際のアーティストを示す

$$\mathcal{L}_d = \sum_{i,j} \omega_i \omega_j |z_i - z_j|$$

二つ目は **Normal Consistency Loss** ( $\mathcal{L}_n$ ) である。2DGS のプリミティブは自身が法線ベクトル  $\mathbf{n}_i =$

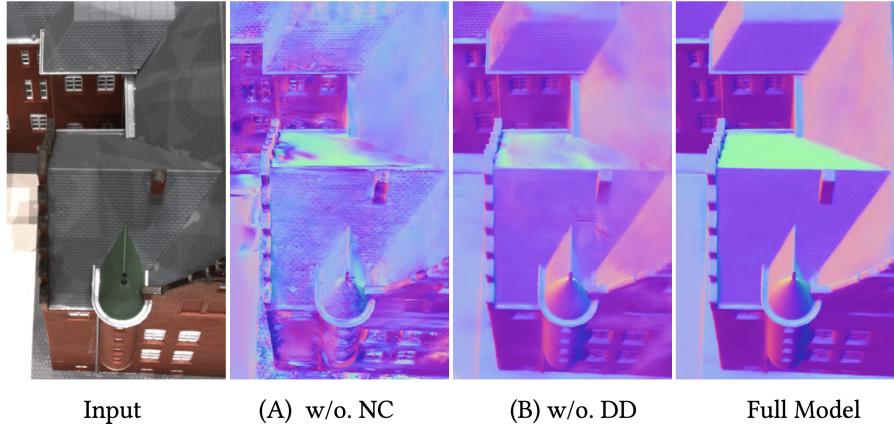


図 3-9: 2DGS の正則化項の効果。左から順に、入力画像、法線整合性正則化なしの表面法線、深度歪み正則化なしの表面法線、そして両方の正則化を適用した 2DGS の結果である。法線整合性正則化を適用しない場合、表面の法線方向にノイズが生じる。一方、深度歪み正則化を省略すると、法線マップがぼやけたものとなる。両方の正則化を併用することで、鋭く滑らかな表面形状を正確に再構成できている。

$\mathbf{t}_u \times \mathbf{t}_v$  を持っている。ここで  $\times$  はベクトルの外積演算を表す。この明示的な法線と、レンダリングされた深度マップから推定される法線  $\mathbf{N}$ （深度の勾配  $\nabla D$  から算出）との整合性を取ることで、滑らかな表面再構成を強制する。

$$\mathcal{L}_n = \sum_i \omega_i (1 - \mathbf{n}_i^\top \mathbf{N})$$

最終的な損失関数は、3DGS の Photometric loss に加え、これらの正則化項を加重和したものである。これらの幾何学的制約により、2DGS は Noise の少ない詳細なメッシュ抽出が可能となり、かつ 3DGS と同等のレンダリング速度と画質を維持している。

メッシュの抽出には TSDF Fusion を用いている。

### 3.4 屈折を扱った関連手法

RefractionNeRF や Refractive-aware MVS、Transparent GS、3DGRT などの諸手法をまとめる。勉強や検証に使えるか。

# 第4章 提案手法

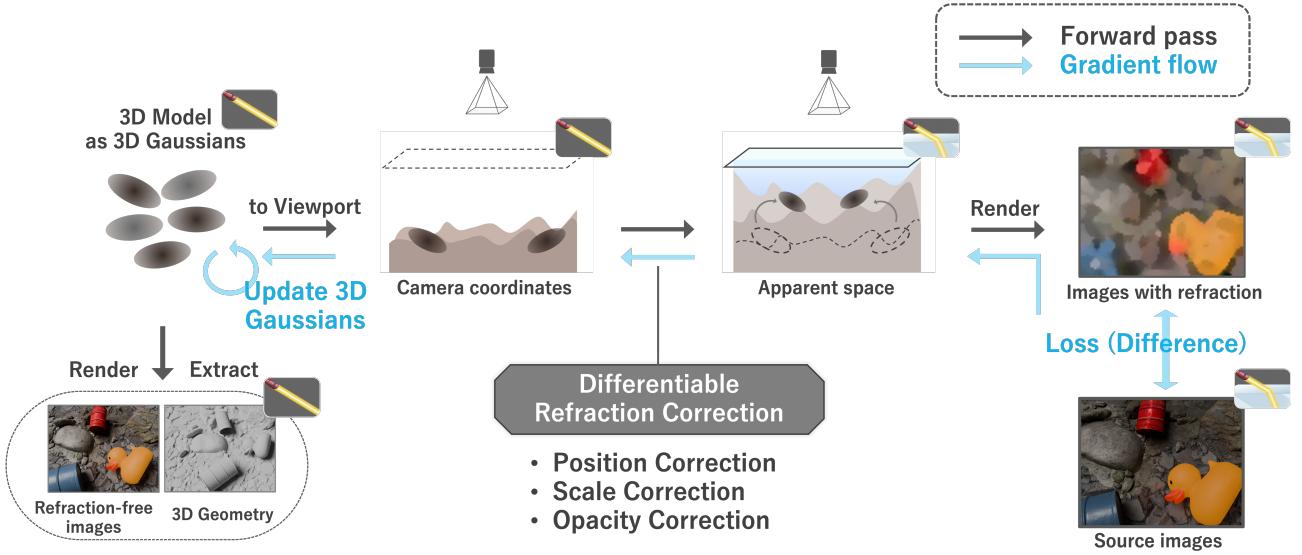


図 4-1: 屈折補正型 Gaussian Splatting の概要。本手法では、3 次元シーンを Gaussian プリミティブの集合で表現し、各カメラ視点ごとに「見かけ空間」へと変換する。具体的には、位置・スケール・不透明度の各パラメータを屈折を考慮した微分可能変換によって補正し、水面における光の屈折現象を物理的にモデル化する。変換後の Gaussian をレンダリングし、屈折による歪みを含む入力画像と比較することで、再構成誤差を逆伝播させて元の Gaussian パラメータを最適化する。これにより、屈折の影響を排除した 3 次元モデルの復元と、フォトリアリスティックな新規視点合成を実現する。

## 4.1 Position Correction (位置補正)

本手法では、空気と水面の境界における光の屈折を考慮し、水中の 3 次元 Gaussian のパラメータを見かけの位置に補正する。

今までの説明から、なぜ屈折を考慮した見た目 (Photometric) を再現することで、屈折の影響を除去した正当な 3 次元形状を復元することが可能となるか。Analysis-by-Synthesis の観点から説明する。

まず、カメラ座標系を定義する。カメラ中心を  $z$  軸上の  $(0, 0, H)$ 、水面を  $xy$  平面 ( $z = 0$ )、水面法線ベクトルを  $\mathbf{n} = (0, 0, 1)$  とする。ここで、水中の任意の Gaussian の中心位置を  $\mathbf{p} = (x, y, z)$  ( $z < 0$ ) とする。水面が平坦かつ静穏であると仮定した場合、屈折現象は視線方向に対して回転対称性を有する。したがって、幾何学的計算を簡略化するため、3 次元座標を動径方向距離  $r = \sqrt{x^2 + y^2}$  を用いて 2 次元の  $rz$  平面へと射影して議論を行う。既存手法 [13·14·16·17] では考慮されていないが、あるカメラ位置からの水中の一点はある見かけの位置に位置に一意にマッピングされる [67·68]。詳細な導出は第 8.1 節を参照。

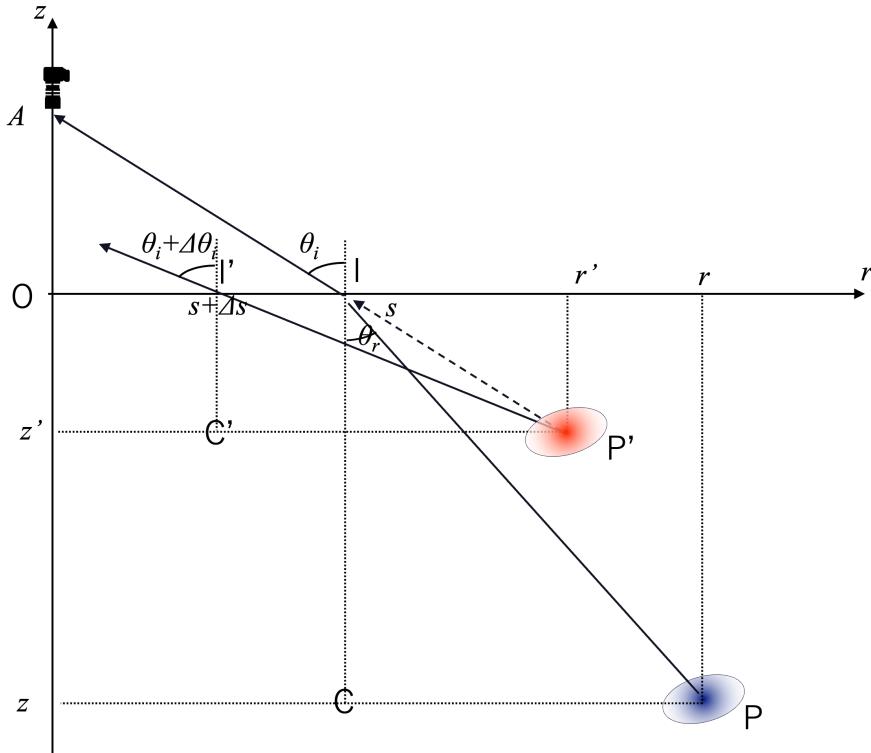


図 4-2: rz 平面における見かけの位置と実際の位置の関係。実際の位置が青色の Gaussian プリミティブで表される場合、見かけの Gaussian は赤色の位置で表される。

屈折を考慮した際、カメラから観測される Gaussian の見かけの中心位置（Apparent position）を  $\mathbf{p}' = (x', y', z')$  とし、rz 平面上での座標を  $(r', z')$  とする。カメラから  $\mathbf{p}'$  へ向かう光線（Ray）と水面との交点を  $(s, 0)$  ( $0 \leq s < r$ ) と定義する。また、入射角（水中）を  $\theta_r$ 、屈折角（空気中）を  $\theta_i$  とし、水の屈折率を  $n$ （空気の屈折率は 1）とすると、スネルの法則より次式が成立する。

どちらを入射角、屈折角と呼ぶか悩ましい。Ray – Tracing 的に視線をベースに考えると空気中が入射角だし、実際の Radiance を考えると水中が入射角である

$$\sin \theta_i = n \sin \theta_r \quad (4.1)$$

幾何学的拘束条件およびスネルの法則より、交点  $s$  は以下の  $s$  に関する 4 次方程式の解  $0 \leq s < r$  として決定される [67]。

$$(1 - n^2)s^4 + 2(n^2 - 1)rs^3 + ((1 - n^2)r^2 - n^2H^2 + z^2)s^2 + 2n^2H^2rs - n^2H^2r^2 = 0$$

ACRS で誤っていた数式を修正済

この方程式の物理的に妥当な解  $s$  は、Newton 法あるいは Ferrari の公式を用いることで数値的に算出可能である。

Implementation は後述だが、収束の速さから Newton を使用

交点  $s$  が求まれば、幾何学的な位置関係およびスネルの法則の微分関係 ( $\triangle IPC, \triangle IP'C', \triangle I'P'C'$  における微小変化) を用いることで、見かけの位置  $(r', z')$  は  $\theta_i$  および  $\theta_r$  の関数として以下の閉形式で導出される。

詳細な導出は Appendix で。ACRS で記述しているのでそれを翻訳

$$\begin{cases} r' = r + (n^2 - 1) \cdot z \cdot \tan^3 \theta_r \\ z' = \frac{1}{n} \cdot \frac{\cos \theta_i^3}{\cos \theta_r^3} \cdot z \end{cases}$$

ここで  $z < 0$  および  $n > 1$ , ( $\theta_i > \theta_r$ ) であることから、 $r' < r$ かつ  $|z'| < |z|$  が常に成立する。これは、水中の物体が屈折の影響により、実位置よりも「浅く」、かつ「光軸（直下）寄りに」観測されるという物理現象と合致する。なお、カメラ直下 (Nadir view:  $r = 0$ ) の特異点においては、上式は  $z' = z/n$  となり、これは従来の航空写真測量等で用いられる近似式 [14] と一致する。

提案手法を 3D Gaussian Splatting の学習プロセスに組み込むためには、この座標変換が微分可能でなければならない。すなわち、実空間上の位置  $p$  に対する見かけの位置  $p'$  の勾配 (Jacobian) が必要となる:

$$J_{app} = \frac{\partial p'}{\partial p} \in \mathbb{R}^{3 \times 3}$$

実装は後述だが、中心差分法でなく、陰関数定理を用いた。そのため、水面高さや屈折率に対しても勾配を流し、屈折率やカメラ位置も最適化対象にすることができる

本研究では、解析的な導出による計算コストと実装の複雑さを回避するため、中心差分法 (Central finite differences) を用いた数値微分により Jacobian を算出する。例えば、 $x$  成分に対する偏微分は次式で近似される。

$$\frac{\partial x'}{\partial x} \approx \frac{p'_x(x + \Delta x) - p'_x(x - \Delta x)}{2\Delta x}$$

この微分可能な「位置補正」により、屈折率と水面高さを与えるだけで、歪んだ観測画像から屈折の影響を除去した正確な 3 次元形状と新視点画像を復元することが可能となる。

## 4.2 3D Covariance $\Sigma^{3D}$ Correction (3D 共分散補正)

3DGS に対する補正と 2DGS に対する補正双方を描く

前節で述べた位置補正  $p \rightarrow p'$  は非線形な変換であり、カメラから見た「見かけの空間 (Apparent space)」は、実空間に対して非一様に歪み、圧縮された空間となる (図 4-3)。特に、入射角が大きい領域や深度が深い領域ほど、光路の屈折による空間圧縮効果は顕著となる。もし、Gaussian の中心位置のみを補正し、その形状 (スケール  $s$  および回転  $R$ ) を実空間のまま補正せずに用いた場合、圧縮された見かけの空間に対して Gaussian が相対的に肥大化して投影されることとなる。

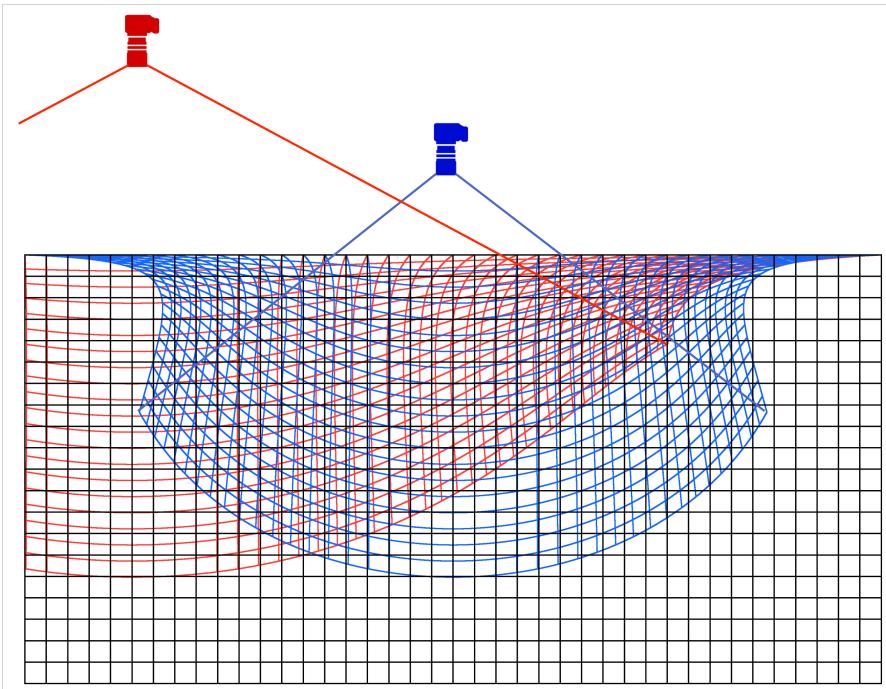


図 4-3: カメラ視点から観測される空間圧縮のイメージ例。カメラ位置に応じて空間が非線形に歪む様子を示している。着色された各グリッドは、カメラ座標系（黒線）での空間格子が屈折効果によってどのように圧縮されるかを表す。

Missing figure

それぞれの Correction を適用した場合、屈折無しで学習したモデルがどのように真値の屈折あり画像に近づいていくか見せる。Lake1 とかを使うといいかも。

これは、レンダリング画像において不自然なぼやけや、境界面の意図しない膨張といった視覚的アティファクトを引き起こす要因となる（図 4-4）。したがって、幾何学的に正確なレンダリングを行うためには、Gaussian の形状を決定する分散共分散行列（Covariance matrix）に対しても、空間の歪みに応じた補正が必要となる。

#### 4.2.1 Covariance Correction for 3DGS

ここで、各 Gaussian プリミティブは十分に小さいと仮定し、その中心  $\mathbf{p}$  近傍における屈折変換を、Jacobian  $J_{app}$ （式 4.1）を用いた局所的な線形近似（Affine 変換）として扱う。

$J_{app}$  の導出を Appendix に載せたい。実装でも結局明示的に算出しているから

実空間における Gaussian の 3 次元分散共分散行列を  $\Sigma^{3D}$  とすると、多変量正規分布の線形変換の性質に基づき、見かけの空間における分散共分散行列  $\Sigma_{app}^{3D}$  は次式で導出される。

$$\Sigma_{app}^{3D} = J_{app} \cdot \Sigma^{3D} \cdot J_{app}^\top$$

この  $\Sigma_{app}^{3D}$  は、屈折によって歪んだ局所空間における Gaussian の適切な形状と配向を表現している。最終的なレンダリングプロセスでは、通常の 3DGS と同様に、この補正された共分散行列  $\Sigma^{3D}_{app}$  に対して射影変換(第 3.3 節)を適用し、画像平面上での 2 次元共分散行列を算出する。これにより、水面屈折特有の非線形な歪みを考慮しつつ、高品質かつアーティファクトの少ない水中シーンの再現が可能となる。

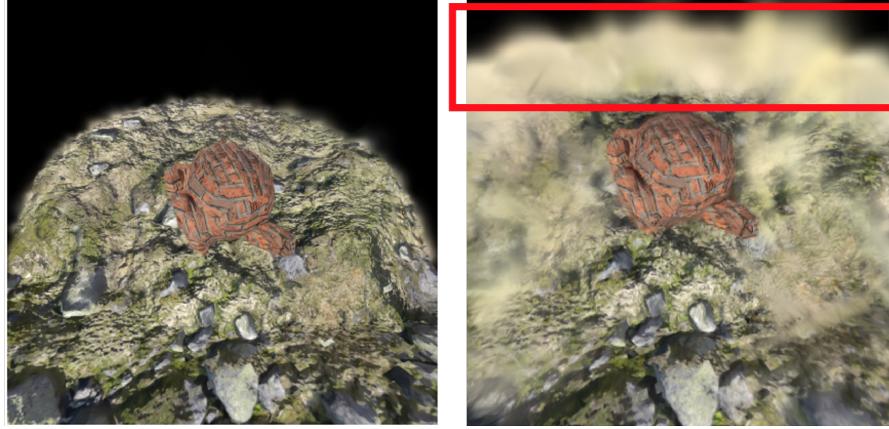


図 4-4: スケール補正を行わない場合のアーティファクトの例。

(左) 屈折面を考慮せず、真値画像で学習した 3DGS モデルのレンダリング結果。

(右) 位置補正のみを適用し、スケール補正を行わない場合のレンダリング結果。赤枠で囲った領域は、入射角が大きい部分を拡大して示している。

#### 4.2.2 Covariance Correction for 2DGS

#### 4.3 Color Correction (色補正)

Color Correction というよりも、「Radiometric Correction」という方が適切でニュアンスが伝わりやすいと思えた

Blender では、どの RGB 補正が使われ画像が出力されるか確認

空気水面間の屈折境界面を通して 3 次元シーンを観測する場合、Gaussian の幾何学的な変動に加え、放射輝度 (Radiance) の変化に伴う「見かけの暗化」が生じる。これは、フレネル反射や水中での散乱・吸収減衰を無視したとしても、幾何光学的な光束  $\Phi$  の広がり (Etendue; エタンデュ) の変化に起因して発生する現象である。Snell の法則 (式式 (4.1)) およびその微分式式 (8.2) より、以下の関係式が導かれる。

$$n^2 \cos \theta_i \sin \theta_i d\theta_i = n^2 \cos \theta_r \sin \theta_r d\theta_r$$

ここで、微小面積要素  $dA$  および方位角の微小変化  $d\psi$  を考慮し、微小立体角  $d\omega = \sin \theta d\theta d\psi$  との関係を用いると、エタンデュ保存則 [69] として知られる以下の関係が得られる。

$$\cos \theta_i d\omega_i dA = n^2 \cos \theta_r d\omega_r dA$$

エタンデュは光学系における保存量であり、境界面での損失がないと仮定した場合、入射光線と屈折光線の間で保存される。したがって、水中 (Radiance  $L_w$ ) から空気中 (Radiance  $L_a$ ) へ光が進む際、光

束の幾何学的变化に伴い、放射輝度は屈折率の二乗の逆数で減衰される：

$$\begin{aligned} L_a &= \frac{d^2\Phi}{dA \cos \theta_i d\theta_i} \\ &= \frac{d^2\Phi}{n^2 dA \cos \theta_r d\theta_r} \\ &= \frac{1}{n^2} L_w \end{aligned}$$

すなわち、空気中のカメラから観測される輝度  $L_a$  は、水中の真の輝度  $L_w$  と比較して  $1/n^2$  倍に減衰し、図 4-5 のようにシーン全体が暗く観測される。

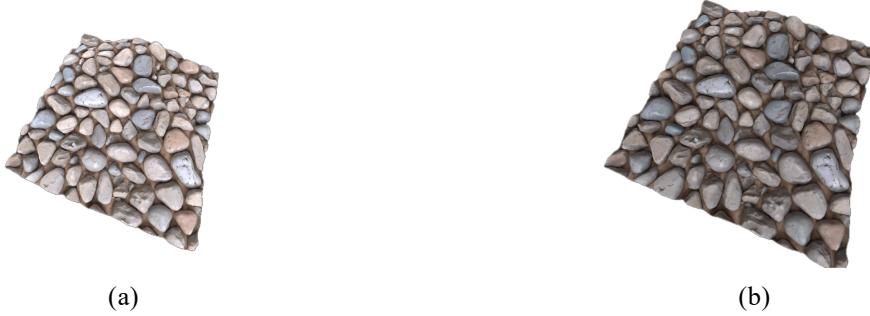


図 4-5: Blender Cycles によるパストレーシングでレンダリングしたシーン例。(a) 屈折効果なし。(b) 水面の屈折を介して観測。(b) は(a)に比べて暗く観測されている。

我々の目的は、この物理的な減衰を補正し、本来の明るさにおける色情報を Gaussian のパラメータとして学習することにある。GSにおいて、シーンの放射輝度は最終的な RGB カラー値にベイクされている。つまり、復元されるシーンの色情報は、撮像記録された画像群における色空間のまま、各 Gaussian の色パラメータ  $c_k$  として保持・レンダリングされる。ここでは、色空間が一般的な sRGB 色空間 (Gamma 補正済み) であると仮定しモデル化を行う。Gamma 値を  $\gamma$  (通常  $\gamma = 2.2$  程度)、RGB から輝度への変換ベクトルを  $\mathbf{I}$  (例  $\mathbf{I}^\top = (0.2126, 0.7152, 0.0722)$ ) とすると、ピクセル強度  $\text{intensity}(\Gamma)$  は以下のようにベクトルの内積で表される：

$$\text{intensity}(\Gamma) = \langle \mathbf{I}, \Gamma^\gamma \rangle$$

ここで  $\Gamma$  は Gamma 補正後のピクセル輝度 (RGB) である。

物理的な放射輝度の関係式 (式 4.3) を満たすためには、補正後の色  $\Gamma'$  は以下の条件を満たす必要がある。

$$\begin{aligned} \text{intensity}(\Gamma'(\mathbf{x})) &= \frac{1}{n^2} \langle \mathbf{I}, \Gamma^\gamma \rangle \\ &= \langle \mathbf{I}, \frac{1}{n^2} \Gamma^\gamma \rangle \end{aligned}$$

これにより、補正後のピクセル値  $\Gamma'$  は以下のように表される:

$$\begin{aligned} (\Gamma')^\gamma &= \frac{1}{n^2} \Gamma^\gamma \\ \Gamma' &= n^{-\frac{2}{\gamma}} \Gamma \end{aligned}$$

すなわち、レンダリング方程式（式 3.3）において、各 Gaussian の色  $c_k$  に対し、次式で表される補正を適用することでエタンデュ保存則を満たすことができる。具体的には、補正後の色  $c'_k$  は

$$c'_k = n^{-\frac{2}{\gamma}} c_k$$

で与えられる。この補正式をレンダリング方程式に代入すると、

$$\begin{aligned} \Gamma'(\mathbf{x}) &= \sum_{k=1}^K c'_k \alpha_k^{\text{pixel}} \prod_{j=1}^{k-1} (1 - \alpha_j^{\text{pixel}}) \\ &= \sum_{k=1}^K \left( n^{-\frac{2}{\gamma}} c_k \right) \alpha_k^{\text{pixel}} \prod_{j=1}^{k-1} (1 - \alpha_j^{\text{pixel}}) \\ &= n^{-\frac{2}{\gamma}} \sum_{k=1}^K c_k \alpha_k^{\text{pixel}} \prod_{j=1}^{k-1} (1 - \alpha_j^{\text{pixel}}) \\ &= n^{-\frac{2}{\gamma}} \Gamma(\mathbf{x}) \end{aligned}$$

となり、第 4.3 節で示された補正を満たすことで、エタンデュ保存則から導かれる屈折による画像の暗化を相殺し、水中シーン本来の色情報を復元可能である。ただし、本稿ではフレネル反射や水中における散乱・吸収減衰の効果は考慮していない。これらの詳細な光学効果については今後の課題とする。

# 第5章 Dataset

## 5.1 Simulated Dataset

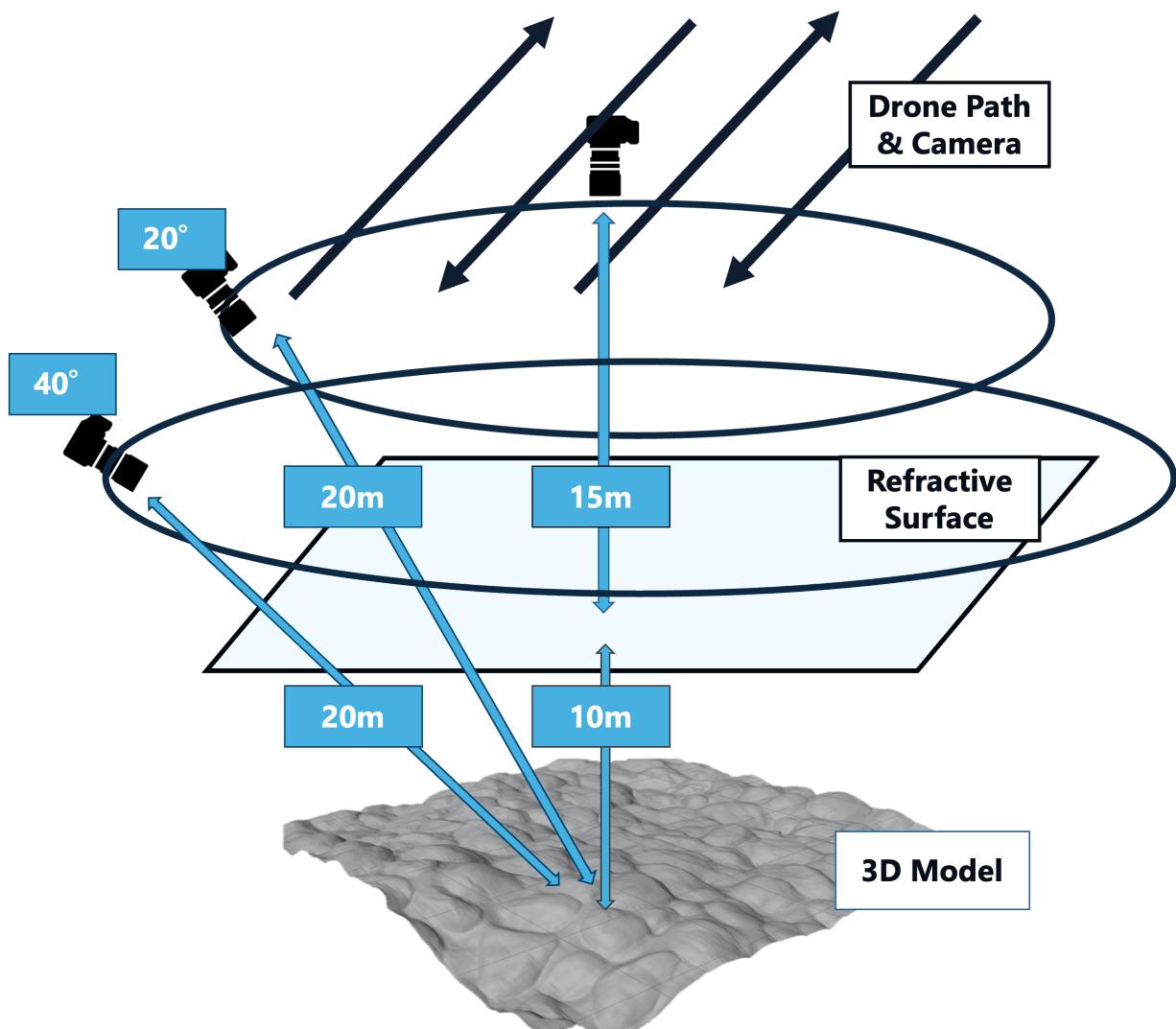


図 5-1: データ取得環境の概要。直下視 (Nadir) に加え、斜視 (Oblique View) の 2 種類の起動から対象の地点を撮影した。

\* 実際の取得画像を示すべき\* 新たな Blender データセットを完成させ、10 種類ほどの多様なデータセットで複合的な検証を行ったかったが、、、

本節では、本実験で使用したシミュレーションデータセットの詳細について述べる。データ取得環境

の概要を図 5-1 に示す。実験環境の構築にあたり、本研究の主目的である「屈折による幾何学的歪み」の影響を厳密に検証するため、3DCG ソフトウェア Blender[70] を用いたシミュレーション環境を採用した。これにより、実環境の実画像において不可避である水面反射や水中での光の減衰・散乱の影響を排除し、純粹な屈折現象のみに焦点を当てた評価を可能とした。河床のシーン構築は、礫のテクスチャに対して Displacement モディファイアを適用することで、不規則な凹凸を持つ微細な地形メッシュを作成した。レンダリングにはパストレーシングエンジンである Blender Cycles を使用し、物理的に忠実な画像を生成した。水面モデルには歪みのない理想的な平面を仮定した。画像の撮影条件は、Gaussian Splatting 等の多視点幾何推定手法に有利な高密度な視差情報を確保するため、オーバーラップ率およびサイドラップ率を共に 85% と高く設定した。撮影軌道は、河床中心を直下視 (Nadir) するグリッド状のパスに加え、天頂角 20° および 40° で対象を取り囲むように円軌道から撮影する条件を採用した (Oblique View)。地形の平均水深は 10 m であり、水面からのカメラ高度も 10 m と、深度に対する相対高度が 1:1 となる非常に屈折の盈虚を受けやすい撮影条件である。このような多角的な観測幾何は、三次元復元における幾何学的制約を強める一方で、図 4-3 に示すように、浅い角度での観測に伴う大きな屈折歪みや見かけ位置のズレを導入することにも繋がる。センサモデルには歪みのないピンホールカメラモデルを採用し、Field of View (FOV) は 70°、画像解像度は 800 × 800 ピクセルとした。データセットの総枚数は、新規視点合成の評価用画像を除き、計 90 枚である。

# 第6章 Dataset

## 6.1 Field Dataset

### 6.1.1 Study Area

実世界における検証に用いたデータセット (Real-world data) は、日本の京都府・宇治市の宇治川中流域におけるデータセットを使用したデータは、2025年11月15日に二箇所においてフィールド調査にて取得した。取得された UAV images の特徴的な例は、に示している。

実世界における手法の検証には、京都府宇治市を流れる宇治川中流域にて取得したデータセット (Real-world data) を使用した。データの取得は2025年11月15日に実施し、同水域内の異なる2地点を対象とした。

まだ Area3 を使用するか決めていない。2 が環境が異なる (大きな礫が多い河床) だが、RTK 情報がないため厳しい。3 は試したいが、影とそれ以外の輝度差の大きい条件

撮影時の気象条件は快晴であり、気温は約 14°C、湿度は約 70% と、光学センサーによる観測に非常に適した環境であった。

本調査では、流速が速い本流ではなく、分派する、または地下から湧き出て水が滞留している副次的な水路を対象とした。この水域は波が穏やかで透明度が極めて高く、水底のテクスチャが視認しやすいという特徴を持つ。目視による推定最大水深は約 1.5m であり、水面の一部には浮草が確認された。取得された UAV 画像の特徴的な例は、に示す。

\* ここでは、撮影エリアでの特徴に終始するぞ。\* 水面には水草が浮いているぞ。\* 本流ではなく、水たまりになっている Second Channel において撮影したぞ。\* 最大深度は 1.5m ほどだと推定されるぞ。\* Texture も比較的豊富だぞ。\* 透明度が高く、波も穏やかだぞ。\* 気温や湿度、天気はこの程度だった。(撮影時 14 度、湿度 60%、快晴)(後に正確に調べ直す)

Missing  
figure

撮影エリアの概要 GIS を使って、論文らしい格好いい図を作る。1 だけではなく、3 でもデータを作りたいから、こだわるのはまた後で。

### 6.1.2 UAV Platform



画像の取得には、RTK-GPS (Real-Time Kinematic Global Positioning System) を搭載した UAV (Unmanned Aerial Vehicle) プラットフォームを使用した。RTK 測位を用いることで、センチメートル級の高精度な位置情報が各画像に付与されるため、地上基準点 (GCP: Ground Control Points) を現地に設置することなく、高精度な SfM (Structure from Motion) 処理が可能となっている。

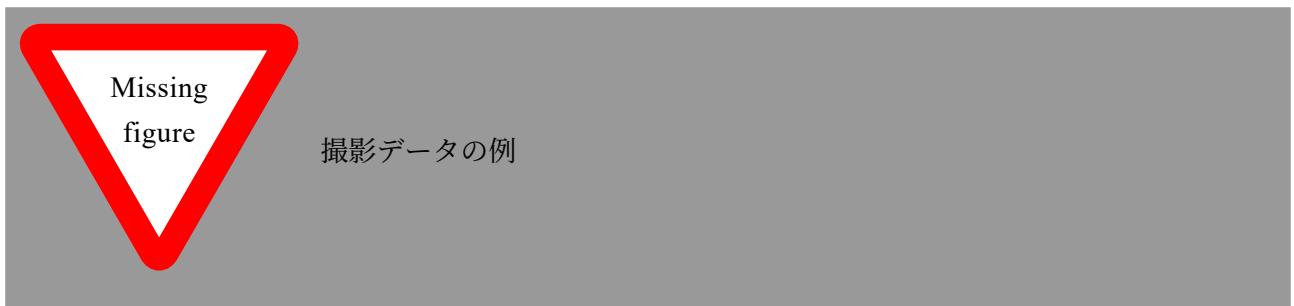
根拠となる論文を引っ張ってくる

また、本研究では水底の三次元形状およびテクスチャを鮮明に捉えることを最優先としたため、カメラレンズに円偏光 (Circular Polarized Light :CPL) フィルターを装着した。これにより、水面における太陽光の反射 (Sun Glint) を物理的に抑制している。偏光フィルターの使用は取得画像の放射輝度値 (Radiance) に影響を与え、厳密なラジオメトリー補正の不確実性を増加させる可能性があるが、本研究の範囲においては水面反射に伴う環境光のモデル化を考慮していないため、水面下の可視性を最大化するアプローチを採用了。

表 6-1: UAV プラットフォームおよびカメラの主要諸元

| Category                        | Specification                    |
|---------------------------------|----------------------------------|
| <b>UAV Platform</b>             |                                  |
| Model                           | DJI Mavic 3 Enterprise           |
| Positioning System              | Onboard RTK Module               |
| <b>Camera Specifications</b>    |                                  |
| Image Sensor                    | 4/3 CMOS                         |
| Effective Pixels                | 20 MP                            |
| Max Image Size                  | 5280 × 3956 pixels               |
| Field of View (FOV)             | 84°                              |
| Equivalent Focal Length         | 24 mm                            |
| Aperture                        | f/2.8                            |
| Focus Range                     | 1 m to $\infty$                  |
| Additional Filter               | Circular Polarizing (CPL) Filter |
| <b>RTK Positioning Accuracy</b> |                                  |
| Horizontal (RMS)                | 1 cm + 1 ppm                     |
| Vertical (RMS)                  | 1.5 cm + 1 ppm                   |

### 6.1.3 Data Acquisition



## 第7章 Experiments

## 7.1 Workflow

本節では、実環境データを対象にした、屈折を考慮した水中3次元再構成のための完全なパイプラインを概説する。

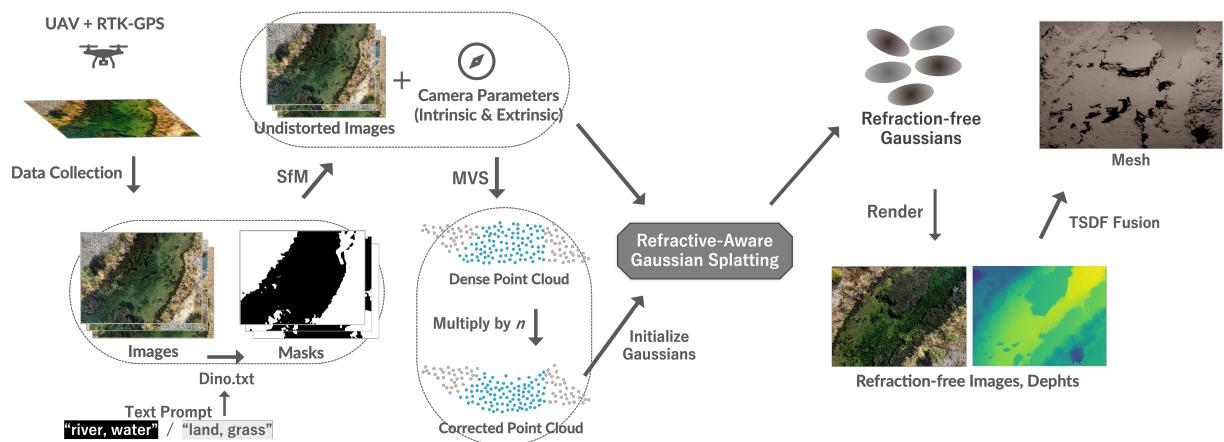


図 7-1: 実環境データを対象にした、屈折を考慮した水中 3 次元再構成のための完全なパイプライン。

### 7.1.1 データ収集とセマンティック・セグメンテーション

## UAVによる画像取得とジオリファレンス

RTK-GPS (Real-Time Kinematic GPS) を搭載した UAV を用いて、河川および陸地を含む対象領域を撮影する。各画像には、RTK 測位による高精度なカメラ外部パラメータ（緯度・経度・高度）がメタデータとして付与される。これにより、SfM において拘束条件を追加し、実スケールでの再構成精度を保証する。

Field Study の具体的な詳細は前章の Dataset で

## Vision-Language Modelによるマスク生成

RA-GS は、入力として正確なカメラ内部・外部パラメータを期待する。これらは既存のパイプラインと同様 [20-49] に、SfM を用いて推定する。屈折面が存在することで、SfM 前提の光の直進性が成立しなくなり、精度は悪化し、また、屈折の影響の大きいシーンでは特徴点の RANSAC による整合ができなくなる場合も考えられる。そこで、SfM に入力画像とその除外エリアを示すマスク画像を入力する。これは、一般的のシーンの三次元復元でも一般的な手法であり、陸域であれば、人物や車両などの移動体、

鏡などの反射物を除外する。本研究での結果は商用フォトグラメトリソフトである RealityScan[71]を使用したが、COLMAP や他のフォトグラメトリソフトウェアでも同様の手法が実現可能である。

この手法では、画像の取得で常に十分な陸域を含める必要があり、UAV の高度や角度に関しての柔軟性を失う。後述するように、RA-GS では、カメラ外部パラメータに関しての勾配も追跡可能で、Photometric な情報からのカメラポーズ推定も可能である。外部パラメータの初期値には GPS によるジオリファレンスとドローンのログデータによる姿勢方角情報を使用する。現在は内部パラメータまでの最適化はできないため、予めカメラはキャリブレーションを行い、歪み補正を行っておく。また、R-SfM[17] を用いて、屈折領域込みでのカメラポーズ推定も可能である。

現在 Code も公開されているが、正直いい実装には思えなかった。

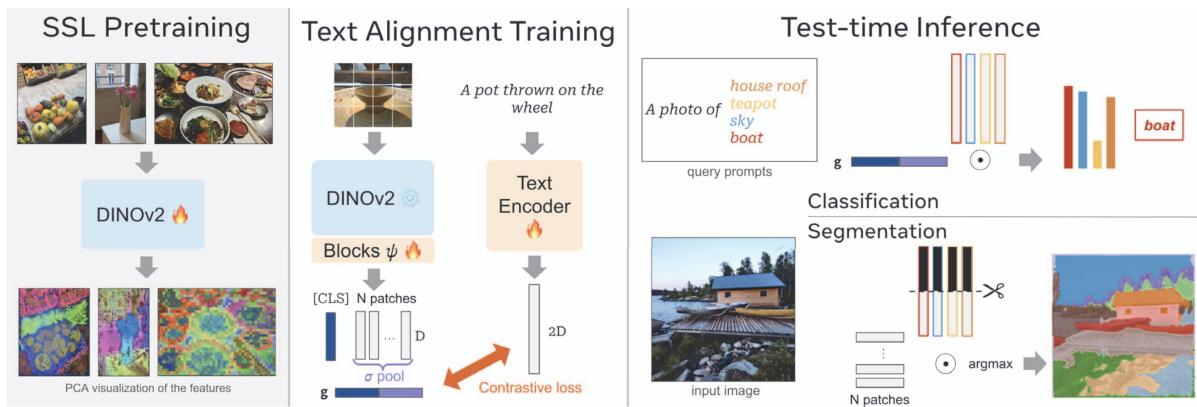


図 7-2: Dino.txt の概要。[\[72\]](#) から引用。(左):DinoV2[\[73\]](#) による自己教師あり学習(Self-Supervised Learning: SSL)による特徴量をPCAで可視化したものである。(本研究では、DinoV3[\[74\]](#) を使用している)(中央):は学習済 SSL と、ゼロから学習したテキストエンコーダを整合させるトレーニング戦略を示す。視覚エンコーダ上に軽量なビジョンブロックを追加することで、テキストとの整合性をさらに向上させている。モデル全体はわずか 5 万イテレーションで学習可能であり、ゼロショット分類およびオープン語彙セグメンテーションの両タスクにおいて最先端性能を達成した(右):出力結果例。(入力画像と、それに対応する zero-shot 分類や open-vocabulary セグメンテーションの結果)。

Missing figure

実際のマスク作成例。定量も含めて示す

本研究の浅水域を対象とした、水域と陸域を区別する。前章で示した宇治川のデータを例に上げると、水面はマスクにより除外し(0 値)、陸域には陸地や草木などが含まれ、それらを非マスクとして残す(1 値)。今回は非マスクとして扱ったが、風の強い日では草木は揺らめくことで幾何推定の精度が悪化するため、その場合はマスクとして除外する必要があるし、宇治川データセットでは水面に多数の浮草が凝集して存在し、それらは屈折の影響を受けないため、SfM による幾何推定の対象に含めたいため非マスクとして抽出したい。このように、Automated: Manual labeling for hundreds images is time consuming ! Robustness: robustly across diverse environments. Zero-Shot: Fine-tuning models for each

new environment is also time consuming impractical ! Controllability : flexible specification using natural language. “I want to remove those ‘ship and shadow’ !” のような条件が必要。

Automated: Manual labeling for hundreds images is time consuming ! Robustness: robustly across diverse environments. Zero-Shot: Fine-tuning models for each new environment is also time consuming impractical ! Controllability : flexible specification using natural language. “I want to remove those ‘ship and shadow’ ! ”

本研究では、オープン語彙物体検出モデル（DINO.txt[[72](#)[74](#)]）を採用し、テキストプロンプトによるゼロショット・セグメンテーションを行う。

Dino の説明は、40prelim の Forward 3D Reconstruction でシたいと思ったがしないかも

具体的には、”river, water” および ”land, grass” というテキストプロンプトからノイズを除去した二値マスク（Binary Masks）を生成する。

# 第8章 APPENDIX

## 8.1 見かけの深度 (Apparent Depth) の導出

カメラ座標系においてカメラ位置を  $z$  軸上  $(0, 0, H)$  に、また水面を平坦な  $xy$  平面（法線ベクトルが一定、すなわち静水面とする）に配置する。水中に存在する各 3 次元 Gaussian プリミティブの位置  $\mathbf{p}$  は  $(x, y, z)$  (ただし  $z < 0$ ) で表される。簡単化のため、問題を 2 次元化し、位置を  $r = \sqrt{x^2 + y^2}$  として  $rz$ -平面上に射影する（図 4-2 を参照）。

幾何光学原理およびスネルの法則 (Snell's law) より、水中点  $\mathbf{p}$  を観測するカメラ視点から見た「見かけの位置」は、観測方向に依存せず定まる [67-68]。ここではその詳細な導出を示す。

Gaussian 中心の見かけの位置を  $\mathbf{p}'(r', z')$  とする。カメラ位置と見かけの位置  $\mathbf{p}'$  を結ぶ入射線と水面との交点を  $(s, 0)$  ( $0 < s < r$ ) とおく。

スネルの法則 (空気屈折率 1、水屈折率  $n$ 、入射角  $\theta_i$ 、屈折角  $\theta_r$ ) は次式で表される。

$$\sin \theta_i = n \sin \theta_r$$

各三角関数項は、カメラ、交点、原点などで構成される三角形  $\triangle OAI$ ,  $\triangle IPC$  により幾何的に表現できる。

上記関係から、 $s$  に関する四次方程式が導かれる。

$$\begin{aligned} (1 - n^2)s^4 + 2(n^2 - 1)rs^3 \\ + ((1 - n^2)r^2 - n^2H^2 + z^2)s^2 \\ + 2n^2H^2rs - n^2H^2r^2 = 0 \end{aligned}$$

(実装ではニュートン法等で数値的に解く。)

わずかなずれ  $\Delta s$  を仮定し、 $s + \Delta s$  の交点、および  $\theta_r + \Delta \theta_r$  の微小変化を考える。三角形  $\triangle IP'C'$  の関係より、

$$-z' \tan \theta_i = r' - s$$

また、 $\triangle I'P'C'$  より

$$\begin{aligned} -z' \tan(\theta_i + \Delta \theta_i) &= (r' - s) + (-\Delta s) \\ -z' (\sin \theta_i + \Delta \theta_i \cos \theta_i) &= (r' - s - \Delta s) (\cos \theta_i - \Delta \theta_i \sin \theta_i) \end{aligned}$$

第 8.1 節と第 8.1 節から  $z'$  を消去し、 $\Delta \rightarrow 0$  の極限をとると

$$r' = s - \sin \theta_i \cos \theta_i \frac{ds}{d\theta_i}$$

が得られる。

また、第 8.1 節と上式より

$$z' = \frac{ds}{d\theta_i} \cos \theta_i^2$$

さらに、三角形  $\triangle IPC$  から

$$r - s = -z \tan \theta_r$$

これを  $\theta_r$  で微分すると

$$\frac{ds}{d\theta_r} = z \cdot \frac{1}{\cos \theta_r^2} \quad (8.1)$$

また、スネルの法則 (4.1) を  $\theta_i$  で微分すると

$$\frac{d\theta_r}{d\theta_i} = \frac{1}{n} \cdot \frac{\cos \theta_i}{\cos \theta_r} \quad (8.2)$$

(8.2),(8.1) より

$$\frac{ds}{d\theta_i} = \frac{z}{n} \cdot \frac{\cos \theta_i}{\cos \theta_r^3}$$

以上をまとめると、見かけの位置  $(r', z')$  はそれぞれ

$$\begin{aligned} & (1 - n^2)s^4 + 2(n^2 - 1)rs^3 \\ & + ((1 - n^2)r^2 - n^2H^2 + z^2)s^2 \\ & + 2n^2H^2rs - n^2H^2r^2 = 0 \end{aligned}$$

と表される。

## 参考文献

- <sup>1</sup>T. Fujii, S. A. Kantoush, Y. Takemon, M. M. Al mamari, and T. Sumi, “Impact of the japanese traditional river training structure “seigyu (crib spur dike)” on river morphology and its geometrical changes”, *Nature-Based Solutions* **6**, 100180 (2024).
- <sup>2</sup>D. Pasquali and A. Marucci, “The effects of urban and economic development on coastal zone management”, *Sustainability* **13**, 6071 (2021).
- <sup>3</sup>J. Thomson, M. Taylor, K. Fryirs, and G. Brierley, “A geomorphological framework for river characterization and habitat assessment”, *Aquatic Conservation: Marine and Freshwater Ecosystems* **11**, 373–389 (2001).
- <sup>4</sup>F. Giordano, G. Mattei, C. Parente, F. Peluso, and R. Santamaria, “Integrating sensors into a marine drone for bathymetric 3d surveys in shallow waters”, *Sensors* **16**, 41 (2015).
- <sup>5</sup>M. Kurowski, J. Thal, R. Damerius, H. Korte, and T. Jeinsch, “Automated survey in very shallow water using an unmanned surface vehicle”, *IFAC-PapersOnLine* **52**, 12th IFAC Conference on Control Applications in Marine Systems, Robotics, and Vehicles CAMS 2019, 146–151 (2019).
- <sup>6</sup>K. Saylam, J. R. Hupp, A. R. Averett, W. F. Gutelius, and B. W. Gelhar, “Airborne lidar bathymetry: assessing quality assurance and quality control methods with leica chiroptera examples”, *International Journal of Remote Sensing* **39**, 2518–2542 (2018).
- <sup>7</sup>J. L. Schönberger and J.-M. Frahm, “Structure-from-motion revisited”, in Conference on computer vision and pattern recognition (cvpr) (2016).
- <sup>8</sup>Y. Furukawa and J. Ponce, “Accurate, dense, and robust multiview stereopsis”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**, 1362–1376 (2010).
- <sup>9</sup>Y. Furukawa and C. Hernández, *Multi-view stereo: a tutorial* (now Publishers Inc, 2015).
- <sup>10</sup>S. P. Bemis, S. Micklethwaite, D. Turner, M. R. James, S. Akciz, S. T. Thiele, and H. A. Bangash, “Ground-based and uav-based photogrammetry: a multi-scale, high-resolution mapping tool for structural geology and paleoseismology”, *Journal of Structural Geology* **69**, 163–178 (2014).
- <sup>11</sup>C. Gomez and H. Purdie, “Uav- based photogrammetry and geocomputing for hazards and disaster risk monitoring – a review”, *Geoenvironmental Disasters* **3**, 10.1186/s40677-016-0060-y (2016).
- <sup>12</sup>J. Igihaut, C. Cabo, S. Puliti, L. Piermattei, J. O’ Connor, and J. Rosette, “Structure from motion photogrammetry in forestry: a review”, *Current Forestry Reports* **5**, 155–168 (2019).
- <sup>13</sup>R. M. Westaway, S. N. Lane, and D. M. Hicks, “Remote sensing of clear-water, shallow, gravel-bed rivers using digital photogrammetry”, *Photogrammetric Engineering and Remote Sensing* **67**, 1271–1282 (2001).
- <sup>14</sup>A. S. Woodget, P. E. Carbonneau, F. Visser, and I. P. Maddock, “Quantifying submerged fluvial topography using hyperspatial resolution uas imagery and structure from motion photogrammetry”, *Earth Surface Processes and Landforms* **40**, 47–64 (2014).
- <sup>15</sup>T. Murase, M. Tanaka, T. Tani, Y. Miyashita, N. Ohkawa, S. Ishiguro, Y. Suzuki, H. Kayanne, and H. Yamano, “A photogrammetric correction procedure for light refraction effects at a two-medium boundary”, *Photogrammetric engineering & remote sensing* **74**, 1129–1136 (2008).
- <sup>16</sup>J. T. Dietrich, “Bathymetric structure-from-motion: extracting shallow stream bathymetry from multi - view stereo photogrammetry”, *Earth Surface Processes and Landforms* **42**, 355–364 (2016).

- <sup>17</sup>A. Makris, V. C. Nicodemou, E. Alevizos, I. Oikonomidis, D. D. Alexakis, and A. Roussos, “Refraction-aware structure from motion for airborne bathymetry”, *Remote Sensing* **16**, 4253 (2024).
- <sup>18</sup>P. Agrafiotis, D. Skarlatos, A. Georgopoulos, and K. Karantzalos, “Shallow water bathymetry mapping from uav imagery based on machine learning”, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **XLII-2/W10**, 9–16 (2019).
- <sup>19</sup>B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: representing scenes as neural radiance fields for view synthesis”, *Communications of the ACM* **65**, 99–106 (2021).
- <sup>20</sup>B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, “3d gaussian splatting for real-time radiance field rendering”, *ACM Transactions on Graphics* **42**, 1–14 (2023).
- <sup>21</sup>日本財団, 日本財団ホームページ, <https://www.nippon-foundation.or.jp/>, Accessed: 2026-01-06.
- <sup>22</sup>GEBCO, *Gebco homepage*, <https://www.gebco.net/>, Accessed: 2026-01-06.
- <sup>23</sup>S. 2030, *Seabed 2030 homepage*, <https://www.seabed2030.org/>, Accessed: 2026-01-06.
- <sup>24</sup>L. Trethewey and 千. 尼丁, *深海の地図をつくる : 五大洋の底をめぐる命がけの競争*, jpn (柏書房, 東京, 2025).
- <sup>25</sup>T. Mason, B. Rainbow, and S. McVey, *Tcolouring the 'white ribbon'*, <https://www.hydro-international.com/content/article/colouring-the-white-ribbon>, Accessed: 2026-01-06, Jan. 2008.
- <sup>26</sup>K. McMahon, M. Berumen, and S. Thorrold, “Linking habitat mosaics and connectivity in a coral reef seascapes”, *Proceedings of the National Academy of Sciences of the United States of America* **109**, 15372–6 (2012).
- <sup>27</sup>国. 水. 河川保全企画室, 河川定期縦横断測量業務実施要領・同解説, [https://www.mlit.go.jp/river/shishin\\_guideline/kasen/pdf/sokuryo\\_youryo.pdf](https://www.mlit.go.jp/river/shishin_guideline/kasen/pdf/sokuryo_youryo.pdf), Accessed: 2026-01-06, Apr. 2018.
- <sup>28</sup>長山学史, 株式会社 第一コンサルタンツ, *I 級河川縦横断測量関連*, <https://www.daiichi-consul.com/works/測量・調査-works/4371>, Accessed: 2026-01-11, Sept. 2017.
- <sup>29</sup>D. S. Research Databases EBooks, *Ecotone*, <https://www.ebsco.com/research-starters/environmental-sciences/ecotone>, Accessed: 2026-01-06.
- <sup>30</sup>A. I. Casalini, P. J. Bouza, and A. J. Bisigato, “Geomorphology, soil and vegetation patterns in an arid ecotone”, *CATENA* **174**, 353–361 (2019).
- <sup>31</sup>C. T. Perry and L. Alvarez-Filip, “Changing geo - ecological functions of coral reefs in the anthropocene”, *Functional Ecology* **33**, edited by N. Graham, 976–988 (2018).
- <sup>32</sup>K. E. Joyce, S. Duce, S. M. Leahy, J. Leon, and S. W. Maier, “Principles and practice of acquiring drone-based image data in marine environments”, *Marine and Freshwater Research* **70**, 95–963 (2018).
- <sup>33</sup>INSIGHTSIAS, *Basic concepts of ecosystem: ecotone*, <https://www.insightsonindia.com/environment/basic-concepts-of-ecosystem/ecotone-ecological-niche/ecotone/>, Accessed: 2026-01-06.
- <sup>34</sup>ア. 株式会社, 空から陸と水深を測る「航空レーザ測深 alb」, [https://www.mlit.go.jp/tec/i-construction/pdf/matching\\_180516\\_siryou\\_1.pdf](https://www.mlit.go.jp/tec/i-construction/pdf/matching_180516_siryou_1.pdf), Accessed: 2026-01-19, May 2018.
- <sup>35</sup>国. 水. 河川保全企画室, 河川管理用三次元データ活用マニュアル(案), [https://www.mlit.go.jp/river/shishin\\_guideline/kasen/pdf/3jigen\\_manual.pdf](https://www.mlit.go.jp/river/shishin_guideline/kasen/pdf/3jigen_manual.pdf), Accessed: 2026-01-11, Feb. 2020.
- <sup>36</sup>B. Janga, G. P. Asamani, Z. Sun, and N. Cristea, “A review of practical ai for remote sensing in earth sciences”, *Remote Sensing* **15**, 10.3390/rs15164112 (2023).
- <sup>37</sup>J. He, S. Zhang, X. Cui, and W. Feng, “Remote sensing for shallow bathymetry: a systematic review”, *Earth-Science Reviews* **258**, 104957 (2024).

- <sup>38</sup>P. Kujawa and F. Remondino, “A review of image- and lidar-based mapping of shallow water scenarios”, *Remote Sensing* **17**, 10.3390/rs17122086 (2025).
- <sup>39</sup>P. Agrafiotis, K. Karantzalos, A. Georgopoulos, and D. Skarlatos, “Correcting image refraction: towards accurate aerial image-based bathymetry mapping in shallow waters”, *Remote Sensing* **12**, 10.3390/rs12020322 (2020).
- <sup>40</sup>G. Mandlburger, M. Kölle, H. Nübel, and U. Soergel, “Bathynet: a deep neural network for water depth mapping from multispectral aerial images”, *PFG - Journal of Photogrammetry, Remote Sensing and Geoinformation Science* **89**, 71–89 (2021).
- <sup>41</sup>Y. Zhan, S. Nobuhara, K. Nishino, and Y. Zheng, “Nerfrac: neural radiance fields through refractive surface”, in *2023 ieee/cvf international conference on computer vision (iccv)* (Oct. 2023), pp. 18356–18366.
- <sup>42</sup>C. Tomasi and T. Kanade, “Shape and motion from image streams under orthography: a factorization method”, *International Journal of Computer Vision* **9**, 137–154 (1992).
- <sup>43</sup>R. Hartley and A. Zisserman, in *Multiple view geometry in computer vision* (Cambridge University Press, Mar. 2004).
- <sup>44</sup>D. G. Lowe, “Distinctive image features from scale-invariant keypoints”, *International Journal of Computer Vision* **60**, 91–110 (2004).
- <sup>45</sup>R. Hartley, “In defense of the eight-point algorithm”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**, 580–593 (1997).
- <sup>46</sup>M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography”, in *Readings in computer vision* (Elsevier, 1987), pp. 726–740.
- <sup>47</sup>B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, “Bundle adjustment — a modern synthesis”, in *Vision algorithms: theory and practice* (Springer Berlin Heidelberg, 2000), pp. 298–372.
- <sup>48</sup>C. Sweeney, *Theia multiview geometry library: tutorial & reference*, <http://theia-sfm.org>, Accessed: 2025-12-18.
- <sup>49</sup>J. L. Schönberger, E. Zheng, J.-M. Frahm, and M. Pollefeys, “Pixelwise view selection for unstructured multi-view stereo”, in *Computer vision - eccv 2016* (Springer International Publishing, 2016), pp. 501–518.
- <sup>50</sup>M. Bleyer, C. Rhemann, and C. Rother, “Patchmatch stereo - stereo matching with slanted support windows”, in *Proceedings of the british machine vision conference 2011*, BMVC 2011 (2011), pp. 14.1–14.11.
- <sup>51</sup>Y. Liao, Y. Di, H. Zhou, K. Zhu, M. Lu, Q. Duan, and J. Liu, “A survey on neural radiance fields”, *ACM Comput. Surv.* **58**, 10.1145/3758085 (2025).
- <sup>52</sup>G. Chen and W. Wang, “A survey on 3d gaussian splatting”, *ArXiv abs/2401.03890* (2024).
- <sup>53</sup>C. Yan, D. Qu, D. Xu, B. Zhao, Z. Wang, D. Wang, and X. Li, “Gs-slam: dense visual slam with 3d gaussian splatting”, in *2024 ieee/cvf conference on computer vision and pattern recognition (cvpr)* (June 2024), pp. 19595–19604.
- <sup>54</sup>J. Zheng, Z. Zhu, V. Bieri, M. Pollefeys, S. Peng, and A. Iro, “Wildgs-slam: monocular gaussian splatting slam in dynamic environments”, in *Proceedings of the ieee/cvf conference on computer vision and pattern recognition (cvpr)* (2025).
- <sup>55</sup>H. Matsuki, R. Murai, P. H. J. Kelly, and A. J. Davison, “Gaussian splatting slam”, in *2024 ieee/cvf conference on computer vision and pattern recognition (cvpr)* (June 2024), pp. 18039–18048.
- <sup>56</sup>A. Moreau, J. Song, H. Dhamo, R. Shaw, Y. Zhou, and E. Pérez-Pellitero, “Human gaussian splatting: real-time rendering of animatable avatars”, in *2024 ieee/cvf conference on computer vision and pattern recognition (cvpr)* (June 2024), pp. 788–798.

- <sup>57</sup>Z. Shao, Z. Wang, Z. Li, D. Wang, X. Lin, Y. Zhang, M. Fan, and Z. Wang, “Splattingavatar: realistic real-time human avatars with mesh-embedded gaussian splatting”, in *2024 ieee/cvf conference on computer vision and pattern recognition (cvpr)* (June 2024), pp. 1606–1616.
- <sup>58</sup>Y. Chen, H. Xu, C. Zheng, B. Zhuang, M. Pollefeys, A. Geiger, T.-J. Cham, and J. Cai, “Mvsplat: efficient 3d gaussian splatting from sparse multi-view images”, in *Computer vision – eccv 2024* (Springer Nature Switzerland, Oct. 2024), pp. 370–386.
- <sup>59</sup>L. S. Aira, G. Facciolo, and T. Ehret, “Gaussian splatting for efficient satellite image photogrammetry”, in *2025 ieee/cvf conference on computer vision and pattern recognition (cvpr)* (June 2025), pp. 5959–5969.
- <sup>60</sup>H. Li, W. Song, T. Xu, A. Elsig, and J. Kulhanek, “WaterSplatting: fast underwater 3D scene reconstruction using gaussian splatting”, *3DV* (2025).
- <sup>61</sup>M. Zwicker, H. Pfister, J. van Baar, and M. Gross, “Ewa volume splatting”, in *Proceedings visualization, 2001. vis ’01.* (2001), pp. 29–538.
- <sup>62</sup>Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity”, *IEEE Transactions on Image Processing* **13**, 600–612 (2004).
- <sup>63</sup>D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization”, arXiv preprint arXiv:1412.6980 (2014).
- <sup>64</sup>G. Grubert, F. Barthel, A. Hilsmann, and P. Eisert, “Improving adaptive density control for 3d gaussian splatting”, in *Proceedings of the 20th international joint conference on computer vision, imaging and computer graphics theory and applications* (2025), pp. 610–621.
- <sup>65</sup>J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, “Mip-nerf 360: unbounded anti-aliased neural radiance fields”, in *2022 ieee/cvf conference on computer vision and pattern recognition (cvpr)* (June 2022).
- <sup>66</sup>B. Huang, Z. Yu, A. Chen, A. Geiger, and S. Gao, “2d gaussian splatting for geometrically accurate radiance fields”, in *Special interest group on computer graphics and interactive techniques conference conference papers, SIGGRAPH ’ 24* (July 2024), pp. 1–11.
- <sup>67</sup>A. B. Nassar, “Apparent depth.”, *Physics Teacher* **32**, 526–29 (1994).
- <sup>68</sup>G. T. Missailidis and I. Guilhon, “Determination of the position of an image formed by flat refracting surfaces at large angles of incidence: three different didactic approaches”, *Revista Brasileira de Ensino de Física* **47**, 10.1590/1806-9126-rbef-2024-0417 (2025).
- <sup>69</sup>R. Winston, L. Jiang, and M. Ricketts, “Nonimaging optics: a tutorial”, *Advances in Optics and Photonics* **10**, 484 (2018).
- <sup>70</sup>B. O. Community, *Blender - a 3d modelling and rendering package*, Blender Foundation (Stichting Blender Foundation, Amsterdam, 2025).
- <sup>71</sup>*Realityscan: a photogrammetric system for 3d reconstruction.*
- <sup>72</sup>C. Jose, T. Moutakanni, D. Kang, F. Baldassarre, T. Darcet, H. Xu, D. Li, M. Szafraniec, M. Ramamonjisoa, M. Oquab, O. Siméoni, H. V. Vo, P. Labatut, and P. Bojanowski, “Dinov2 meets text: a unified framework for image- and pixel-level vision-language alignment”, in *2025 ieee/cvf conference on computer vision and pattern recognition (cvpr)* (June 2025), pp. 24905–24916.
- <sup>73</sup>M. Oquab, T. Darcet, T. Moutakanni, H. V. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, R. Howes, P.-Y. Huang, H. Xu, V. Sharma, S.-W. Li, W. Galuba, M. Rabbat, M. Assran, N. Ballas, G. Synnaeve, I. Misra, H. Jegou, J. Mairal, P. Labatut, A. Joulin, and P. Bojanowski, *Dinov2: learning robust visual features without supervision*, 2023.
- <sup>74</sup>O. Siméoni, H. V. Vo, M. Seitzer, F. Baldassarre, M. Oquab, C. Jose, V. Khalidov, M. Szafraniec, S. Yi, M. Ramamonjisoa, F. Massa, D. Haziza, L. Wehrstedt, J. Wang, T. Darcet, T. Moutakanni, L. Sentana, C. Roberts, A. Vedaldi, J. Tolan, J. Brandt, C. Couprie, J. Mairal, H. Jegou, P. Labatut, and P. Bojanowski, *DINOV3*, 2025.